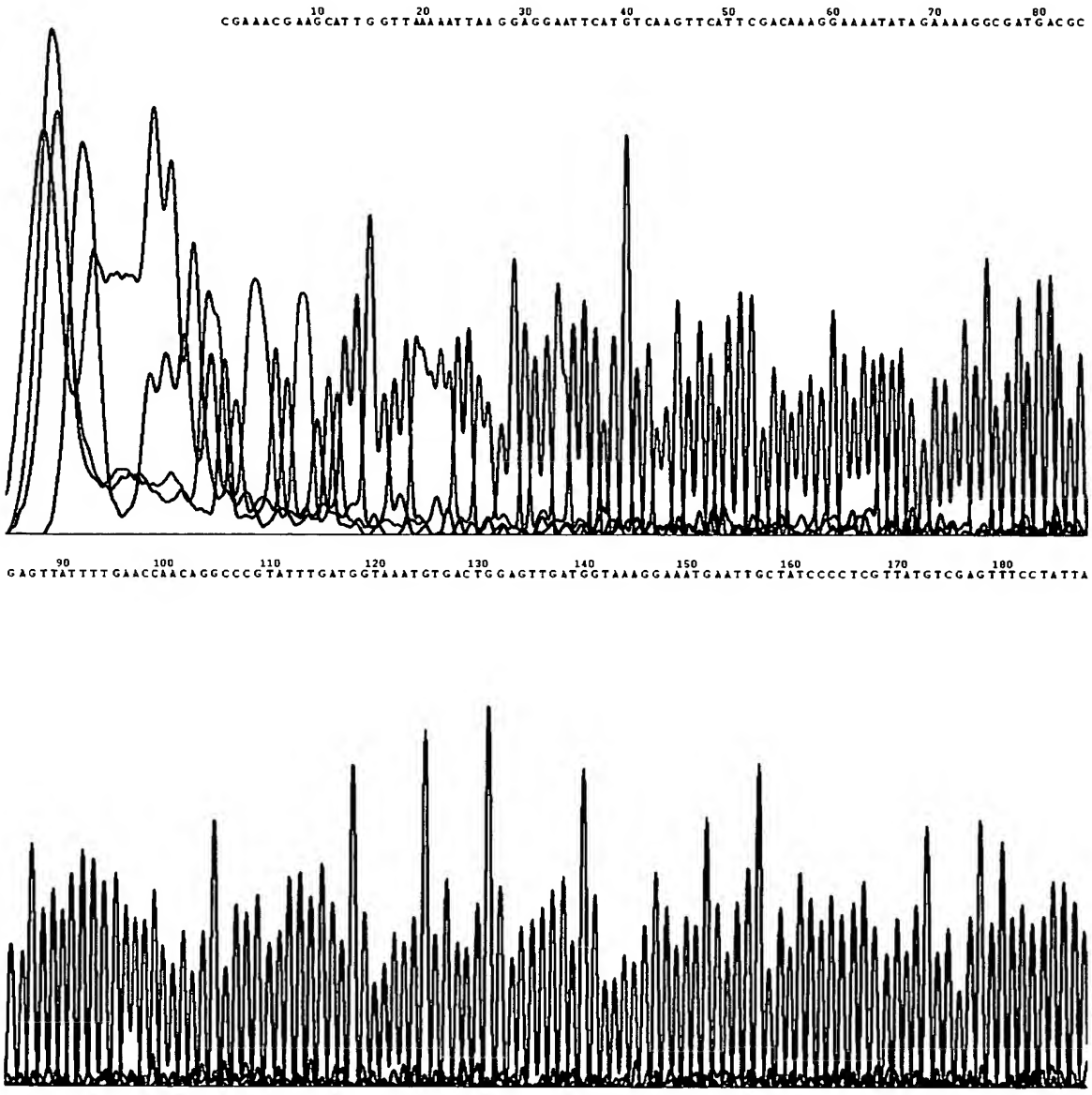
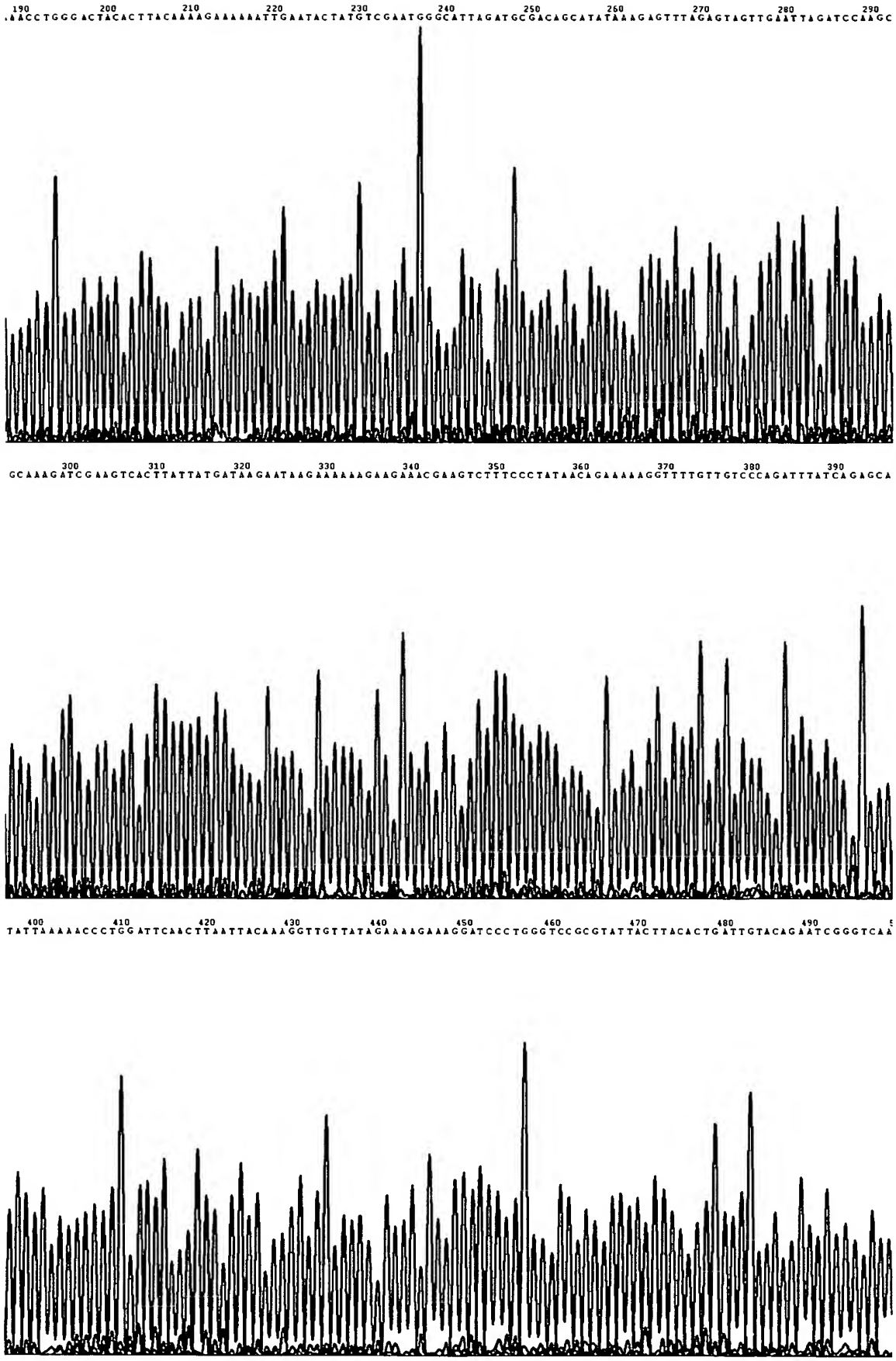
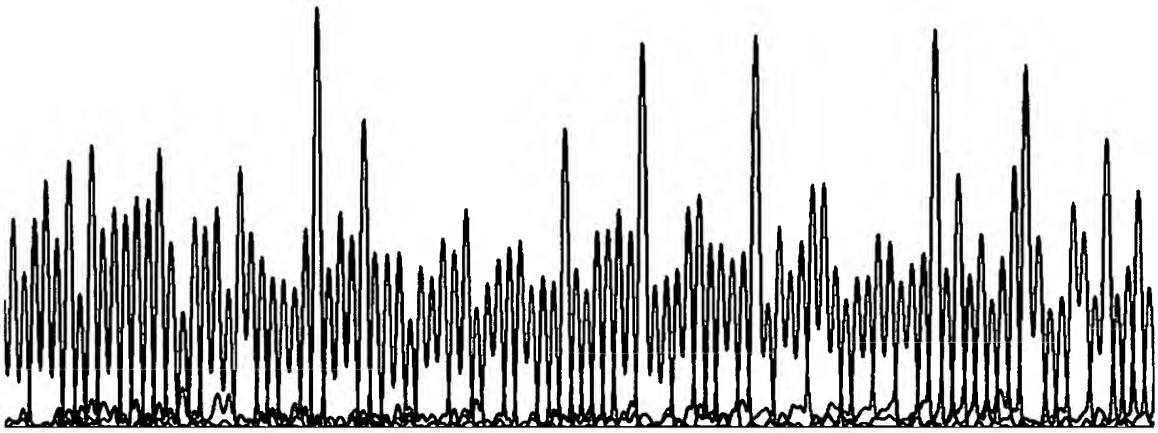
The results of sequencing for the nucleotide sequence for the fusion protein STH. The sequence for gene of interest is located from nucleotide 41 to nucleotide 664; and the sequence for linker peptide is located from nucleotide 449 to nucleotide 466.

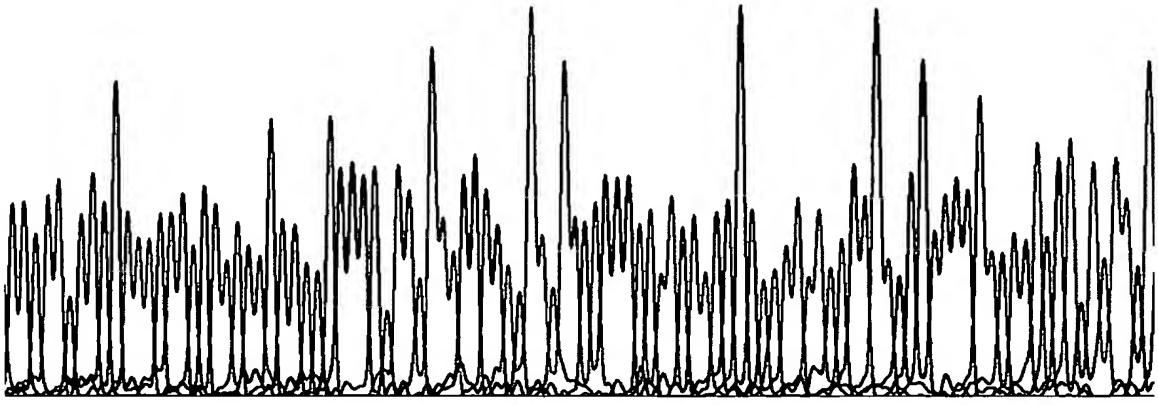



190 200 210 220 230 240 250 260 270 280 290
AACCTGGGACTACACTTACAAAAGAAAAAATTGAATACTATGTCGAATGGGCATTAGATGCGACAGCATATAAAGAGTTTAGAGTAGTTGAATTAGATCCAAGC
300 310 320 330 340 350 360 370 380 390
GCAAAGATCGAAGTCACTTATTATGATAAGAATAAGAAAAAAGAAGAAACGAAGTCTTTCCCTATAACAGAAAAAGGTTTTGTTGTCCCAGATTTATCAGAGCA
400 410 420 430 440 450 460 470 480 490
TATTAAAAACCCTGGATTCAACTTAATTACAAAGGTTGTTATAGAAAAGAAAGGATCCCTGGGTCCGCGTATTACTTACACTGATTGTACAGAATCGGGTCAA

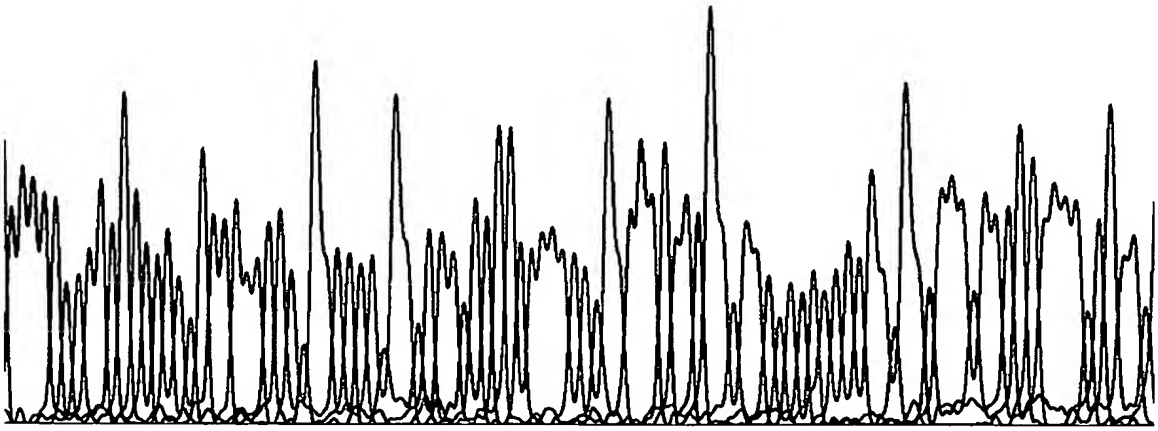
500 510 520 530 540 550 560 570 580 590 600
AATTTGTGCCTCTGCGAGGGAAGCAATGTTTGCGGTAAAGGCAATAAGTGCATATTGGGTTCTAATGGAAAGGGCAACCAATGTGTCACTGGCGAAGGTACA



610 620 630 640 650 660 670 680 690 700
CCGAAGCCTGAAAGCCATAACAACGGCGATTTCGAAGAAATTCCAGAAGAATATTTACAATAACTGCAGCCAAGCTTGGCTGTTTTGGCGGATGAGAGAAGA



710 720 730 740 750 760 770 780 790 800
TTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCG



The nucleotide sequence for the fusion protein STH, wherein the nucleotide sequence for the linker peptide is underlined

TCAAGTTCATTCGACAAAGGAAAATATAGAAAAGGCGATGACGCGAGTTAT
TTTGAACCAACAGGCCCGTATTTGATGGTAAATGTGACTGGAGTTGATGG
TAAAGGAAATGAATTGCTATCCCCTCGTTATGTCGAGTTTCCTATTAAACCT
GGGACTACACTTACAAAAGAAAAAATTGAATACTATGTCGAATGGGCATTA
GATGCGACAGCATATAAAGAGTTTAGAGTAGTTGAATTAGATCCAAGCGCA
AAGATCGAAGTCACTTATTATGATAAGAATAAGAAAAAAGAAGAAACGAAG
TCTTTCCCTATAACAGAAAAAGGTTTTGTTGTCCCAGATTTATCAGAGCATA
TTAAAAACCCTGGATTCAACTTAATTACAAAGGTTGTTATAGAAAAGAAAG
GATCCCTGGGTCCGCGTATTACTTACACTGATTGTACAGAATCGGGTCAA
AATTTGTGCCTCTGCGAGGGAAGCAATGTTTGCGGTAAAGGCAATAAGTG
CATATTGGGTTCTAATGGAAAGGGCAACCAATGTGTCACTGGCGAAGGTA
CACCGAAGCCTGAAAGCCATAACAACGGCGATTTCGAAGAAATTCCAGAA
GAATATTTACAATAA

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Protein Expression and Purification 31 (2003) 1–11

Protein Expression & Purification

www.elsevier.com/locate/yprep

Review

A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa

Richard J. Jenny,[a,*] Kenneth G. Mann,[b] and Roger L. Lundblad[c,d]

[a] *Haematologic Technologies, Inc., Essex Junction, VT, USA*
[b] *Department of Biochemistry, University of Vermont, Burlington, VT, USA*
[c] *Department of Pathology, University of North Carolina, Chapel Hill, NC, USA*
[d] *Roger L. Lundblad, LLC, Chapel Hill, NC, USA*

Received 27 February 2003, and in revised form 7 May 2003

Abstract

Expression and purification of proteins in recombinant DNA systems is a powerful and widely used technique. Frequently there is the need to express the protein of interest as a fusion protein or chimeric protein. Fusion protein technology is frequently used to attach a "signal" which can be used for subsequent localization of the protein or a "carrier" which can be used to deliver a "therapeutic" such as a radioactive molecule to a specific site. In addition to these applications, fusion protein technology can be employed for several other useful purposes. Of these, the most frequent reason is to provide a 'tag' or 'handle' which will aid in the purification of the protein. Another useful purpose is to improve the expression or folding of the protein of interest. In these latter two situations, it is often necessary to remove the fusion partner before the recombinant protein of interest can be used for further studies. This removal process involves the insertion of a unique amino acid sequence that is susceptible to cleavage by a highly specific protease. Thrombin and factor Xa are the most frequently used proteases for this application. The purpose of this review is to discuss the application of thrombin and factor Xa for the cleavage of fusion proteins. It is emphasized that while these enzymes are quite specific for cleavage at the inserted cleavage site, proteolysis can frequently occur at other site(s) in the protein of interest. It is necessary to characterize the protein of interest after cleavage from the affinity label to assure that there are no changes in the covalent structure of the protein of interest. Examples are presented which describe the proteolysis of the protein of interest by either factor Xa or thrombin.

The use of recombinant DNA technology for the expression of proteins in both prokaryotic and eukaryotic cells is a well-accepted technology. Often such proteins are expressed as fusion or chimeric proteins. The use of a fusion partner as carrier for either a "signal" or "pharmaceutical" is of significant value [1–6]. A more frequent application is the use of a fusion partner to provide a "tag" which can be used for the subsequent purification. Examples of such "tags" are the hexahistidine [7], streptavidin [8], glutathione-*S*-transferase [9], and maltose-binding protein [10]. These 'tags' are particularly useful because for each there is a corresponding affinity matrix that when used under appropriate conditions will bind and release the 'tag.' Once the chimeric molecule has been isolated, it is usually necessary to remove the "tag" before subsequent use of the expressed protein of interest [11]. The most popular method to remove the "tag" involves the use of a specific protease such as thrombin or factor Xa. This involves the insertion of a unique amino acid sequence that is specific for cleavage by either factor Xa or thrombin between the protein of interest and the fusion partner. In the current article, we review the use of this technology.

Both thrombin and factor Xa are trypsin-like serine proteases which will cleave peptide bonds on the carboxyl side (P'_1) of a basic amino acid residue. While trypsin will effectively cleave peptide bonds when either lysine or arginine precedes the carboxyl group, thrombin

* Corresponding author. Fax: 1-802-878-1776.
E-mail addresses: rjenny@haemtech.net, hti@haemtech.com (R.J. Jenny).

1046-5928/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1046-5928(03)00168-2

and factor Xa have a preference for peptide bonds in which the carboxyl function is provided by an argininyl residue. In addition, while trypsin will cleave essentially all peptide bonds in a protein where the carboxyl moiety is contributed by lysine or arginine, thrombin and factor Xa are far more specific and cleave only certain peptide bonds. This specificity is contributed by the sequence of amino acids surrounding the scissile peptide and more particularly that sequence which precedes the scissile bond. The differences between trypsin and thrombin/factor Xa reflect the fact that trypsin is a "digestive" protease while thrombin and factor Xa are regulatory proteases.

The specificity of thrombin has been studied by a number of investigators [12]. The most common linker sequence is LVPRG or LVPRGS, which is derived from the sequence in bovine factor XIII that is cleaved by thrombin during the activation process [13]. Cleavage occurs at the arginine residue, resulting in the protein of interest being extended at its amino-terminal end by either a G or GS. The linker sequence for factor Xa cleavage is usually IEGR which is derived from the sequences in prothrombin which are hydrolyzed by factor Xa during the conversion of prothrombin to thrombin [14]. While cleavage by thrombin or factor Xa in the linker sequence region is reasonably specific, it is not absolute. Cleavage can occur at other sequence sites in the protein. There have been a limited number of studies on the action of thrombin on proteins other than those involved in blood coagulation. Lonsdale-Eccles et al. [15] studied the hydrolysis of chemically modified proteins by thrombin and factor Xa. Both enzymes hydrolyze peptide bonds in which the carboxyl moiety is contributed by arginine or lysine and hydrolysis at peptide bonds containing *S*-(β-aminoethyl)cysteine is also observed. In performic acid-oxidized bovine chymotrypsinogen, bovine thrombin and factor Xa cleaved at the argininyl residue in the sequence GVYARVTA. Factor Xa also cleaved the following sequence SGLSRIVN which was not cleaved by thrombin. Bovine thrombin cleaved at the arginine residues in the sequence QVRLG in maleated performic acid-oxidized bovine trypsinogen; this sequence was not cleaved by factor Xa. Factor Xa did cleave at the arginine residue in the sequence NSRVA in maleated, performic acid-oxidized bovine trypsinogen; this sequence was also hydrolyzed by thrombin. Thrombin cleaved at the lysine residues in the sequences MKSRNL, RCKPVN, and SSKYPN in performic acid-oxidized ribonuclease. Factor Xa did not hydrolyze any peptide bonds in performic acid-oxidized ribonuclease. The enzymatic hydrolyses were performed in either 0.2 M NH_4CO_3, pH 8.0, or in a pH-stat. The bicarbonate-buffered digest contained the substrate protein at a concentration of 5.0 mg/mL and either factor Xa (200 μg/mL) or thrombin (50 μg/mL). The digest was performed for 8–9 h.

Chang [16] studied the action of thrombin on 30 polypeptide hormones. The sequences cleaved are shown in Table 1. There are significant differences in the rate of hydrolysis of these bonds. With secretin, the time required to obtain 50% cleavage of the Arg–Asp is 11.5 h. After 48 h of hydrolysis, there was limited cleavage of the Arg–Leu peptide bond. It is interesting to note that heparin increased the rate of hydrolysis of the Arg–Asp peptide bond. There was approximately 35% cleavage of the two scissle peptide bonds in vasoactive intestine protein after 48 h. There was 50% cleavage of the Arg–Val in choleocystokinin after 75 min and 23% cleavage of the Lys–Asparagine bond at 24 h. There was 50% cleavage of the Lys–Leu bond at 5 h.

Chang et al. [17] studied the hydrolysis of antibody light chains by thrombin. Mouse κ light chains 10K26 and 10K44 from anti-(azobenzene arsonate) antibodies contain 20 Arg/Lys-X peptide bonds. Only two are cleaved by thrombin; NTLPRTFGG has a 50% cleavage time of 10 min with 0.5 NIH units thrombin and 1 nmol protein in 20 μL of 50 mM NH_4CO_3, pH 8.0, at 25 °C. There is cleavage of the sequence SPIVKSFN, which is hydrolyzed at the lysine residue with a 50% time of 10 h.

The purpose of reviewing the above information was to demonstrate that both thrombin and factor Xa can hydrolyze a variety of peptide bonds in proteins. As such, characterization of the desired product protein after cleavage of the fusion protein is essential. Retention of biological activity does not necessarily imply retention of native covalent structure. The following studies are discussed as examples of "extraneous cleavage" by thrombin or factor Xa in fusion proteins.

Raftery et al. [18] designed a fusion protein between murine S100 protein MRP14 and glutathione-*S*-transferase. The thrombin cleavage sequence LVPRGS was placed between the two proteins. The fusion protein was expressed in *Escherichia coli* and was isolated from the

Table 1
Sequences cleaved by thrombin in polypeptide hormones

Polypeptide homones[a]	Sequence cleaved
Secretin	ELSLSRLRDSA
Secretin	ELSLSRLR (much slower than above)
Vasoactive intestine polypeptide	DNYTRLRK
Vasoactive intestine polypeptide	YTRLRKQM
Choleocystokinin	APSGRVSM
Choleocystokinin	VSMIKNLQ
Dynorphin A	RIRPKLKW
Somatostatin-28	AMAPRERK
Somatostatin-28	NFFWKTFT
Gastrin releasing peptide	KMYPRGNH
Salmon calcitonin	QTYPRTNT

[a] The reaction mixtures contained 0.5 NIH units thrombin and 1.0 nmol peptide in 20 μL of 50 mM NH_4CO_3, pH 8.0, at 25 °C. The conditions were designed to obtain an enzyme/substrate ratio of 1:60 (w/w).

culture media using glutathione–agarose beads. The isolated fusion protein was digested with thrombin (20 NIH units/mL in 25 mM Tris and 250 mM NaCl, pH 7.5, for 60 min at 37 °C). HPLC analysis of the products revealed a component with a molecular mass of 13,062 Da consistent with MRP14 with a GS extension at the amino terminal end. More extensive digestion (3 h) resulted in two components, one with a molecular mass of 13,062 Da and consistent with native MRP14 with GS extension at the NH_2-terminal end and second component with a molecular mass of 11,919 Da. Prolonged digestion (24 h) resulted in only the 11,919 Da protein, corresponding to residues 1–102. This fragment resulted from the thrombin cleavage at the sequence NNPRGHGH near the carboxyl terminus. The truncated MRP14 had reduced zinc binding compared to the native or recombinant full-length MRP14.

Zaitseva et al. [19] studied chimeric proteins, which are components of the "ABC transporters." They prepared a fusion protein between RbsA and RbsC with the factor Xa cleavage sequence IEGRGH between the two proteins. A poly-His leader was attached for the purpose of purification. Digestion with factor Xa in 100 mM sodium phosphate, pH 8.0, with 20% glycerol, 0.1% dodecylmaltoside, 1.5 M NaCl, and 1 mM $CaCl_2$ gave complete digestion of the factor Xa linker sequence but there was also cleavage at site in the RbsC protein at a VSGR (residues 6–9) sequence. The use of 50 mM Tris, pH 7.5, with 20% glycerol, 0.15 M NaCl, 0.01% dodecylmaltoside, and 1 mM $CaCl_2$ resulted in slower hydrolysis, although complete digestion at the factor Xa linker sequence was achieved after 4 h at 23 °C with minimal but detectable cleavage at the secondary site. A factor Xa cleavage site was also inserted between the poly-His sequence and RbsA. This site is hydrolyzed more slowly than the factor Xa cleavage site between RbsA and RbsC.

Muse and Bender [20] studied fusion proteins between the nitrogen assimilation control protein (NAC) from either *Klebsiella aerogenes* (NAC_K) or *E. coli* (NAC_E), and maltose-binding protein. A factor Xa cleavage site IEGR was inserted between the NAC protein and the maltose-binding protein. The fusion proteins were purified by affinity chromatography on amylose–agarose resin. Elution was accomplished with 1 M maltose. Digestion with factor Xa was performed in 100 mM sodium phosphate, pH 7.0, containing 250 mM NaCl, 2.5 mM $MgCl_2$, and 1 mM of 2-mercaptoethanol. One microgram of factor Xa was added to 0.5 mg of the fusion proteins in 100 mL volume at 22 °C for 16 h. The reaction was terminated by the addition of phenylmethylsulfonyl fluoride to a final concentration of 20 mM. A fusion protein with a polyhistidine (His_6) was also prepared. Cleavage of NAC_K-maltose binding protein with factor Xa resulted in the formation of dimers of NAC_K. NAC_E-maltose-binding protein was effectively cleaved by factor Xa but the resulting NAC_E was rapidly degraded (cleavage sites not provided). The NAC_E-maltose-binding protein with the His_6 tail was resistant to cleavage by factor Xa.

Belmouden et al. [21] prepared a fusion protein between rat kidney long-chain L-2-hydroxy acid oxidase and glutathione-*S*-transferase with a factor Xa cleavage site IEGRGIP inserted between the two proteins. Digestion with factor Xa was performed in 50 mM of Tris, 100 mM of NaCl, and 4 mM $CaCl_2$, pH 8.0, at 25 °C. There was little proteolysis detectable (SDS–polyacrylamide gel electrophoresis) at enzyme–substrate ratios of 1:100 or 1:10. Trypsin readily cleaved the factor Xa sequence; there was also proteolysis at other sites resulting in three major bands. It was suggested that the factor Xa cleavage site between the two proteins was inaccessible to factor Xa. Cleavage of the linker site by factor Xa has also been a problem for other investigators. Ko and coworkers [22] prepared a fusion protein between the first nucleotide binding fold (NBF1) of the cystic fibrosis transmembrane conductance regulator (CFTR) and maltose binding protein with a factor Xa-sensitive linker site. The fusion protein resisted cleavage by factor Xa under a variety of solvent conditions; cleavage was accomplished by first incubating the fusion protein in the presence of 0.08% sodium dodecyl sulfate for 16 h at 23 °C, followed by addition of factor Xa at a weight ratio of 1:100 and subsequent reaction at 4 °C for 14 h. Sharma and Rose [23] observed a poor extent of cleavage in a fusion protein comprised of the carboxyl-terminal nucleotide binding domain of *P*-glycoprotein and maltose binding protein. Wang et al. [24] have reported an extremely useful study on fusion proteins between the first nucleotide-binding domain and linker region of human multidrug resistance gene product and glutathione-*S*-transferase, thioredoxin, and maltose-binding protein. Expression of the glutathione-*S*-transferase and thioredoxin fusion constructs as soluble proteins was low while much higher expression of the maltose binding protein construct was observed. Cleavage with factor Xa was unsatisfactory as there was significant non-specific proteolysis; insertion of a thrombin cleavage site resulted in satisfactory product.

The above experiments are cited as case studies of problems associated with the use of thrombin and/or factor Xa to cleave fusion proteins. The point is that there must be careful and thorough structural characterization of the protein of interest after the cleavage step. With appropriate care, cleavage of fusion proteins with either thrombin-sensitive or factor Xa-sensitive sequences can be achieved. A partial compilation of studies using thrombin is given in Table 2 and for factor Xa in Table 3.

An additional issue concerns the purity of the enzymes used in the cleavage reaction. Thrombin, in particular, can be problematic. Contamination with plasmin

Table 2
Thrombin-catalyzed cleavage of fusion proteins

Protein of interest	Fusion protein	Cleavage sequence	Cleavage condition	Reference
ADP-ribosylation factor	GST[a]		50 mM Tris, pH 8.0, with 2.5 mM $CaCl_2$, 6 μg thrombin in 0.5 mL (concentration of fusion protein not given), 4 h at room temperature on a platform rocker	[52]
β_2-Adreno-receptor	His-tag[b]-$G_{S\alpha}$	SRGSLDPRSFLLRN PNDKYEPFWEDEE[c]	100 nM thrombin/1 mg protein per mL in 10 mM Tris–HCl, pH 7.4, at 25 °C for 30 min with constant rotation[d,e]	[53]
Transforming protein E7	GST[a]		100 units thrombin/25 mg fusion protein in 5 mM Hepes–150 mM NaCl, pH 8.0, containing 10 μg $CaCl_2$/mL at 23 °C for 2 h[f]	[54]
CRaf1	His-tag[b]		Conditions not provided	[55][g]
MEK1	GST[a]		Conditions not provided	[55][g]
ERK2	GST[a]		Conditions not provided	[55][g]
Murine osteopontin	GST[a]		0.1 unit thrombin/10 μg fusion protein at 37 ° for 2 h. Solvent conditions not provided	[56]
HTom20	GST[a]		The fusion protein was digested while bound to glutathione–Sepharose 4B as a slurry in 20 mM sodium phosphate, 150 mM NaCl, pH 7.4, with 1 mM dithiothreitol. Fifty micrograms of thrombin was added and digestion was allowed to proceed for 18 h at 4 °C[h]	[57]
Cystic fibrosis transmembrane conductance regulator components	GST[a]		Digestion of the column-bound by fusion protein by thrombin. Five units thrombin/100 μg protein in phosphate-buffered saline, pH 7.3, at 25 °C for 2 h	[58]
Chlorocatechol 1,2-dioxygenase	His-tag[b]		1 unit thrombin/mg fusion protein in 20 mM Tris–HCl, pH 8.0, containing 150 mM NaCl and 2.5 mM $CaCl_2$at 18 °C for 90 h	[11]
Tenascin-C and Tenascin-R	GST[a]		Thrombin was used to digest the fusion protein bound to glutathione–Sepharose 4B matrix. Solvent conditions were not provided; Pefabloc SC was included to inhibit residual thrombin	[59]
MRP14	GST[a]	LVPRGS	50 units thrombin/liter culture at 37 °C for 24 h[i]	[18]
IL-6	IgG1 FC	LVPRGS	5 units of thrombin/50 μL phosphate-buffered saline to elute fusion protein bound to Protein A–Sepharose	[60]
QPs3	GST[a]		1 μg thrombin/500 μg fusion protein in 50 mM Tris–HCl, pH 8.0 containing 0.25 M sucrose. Time/temperature not provided[f]	[61]
Heavy chain of tetanus toxin	MBP[j]		1 unit thrombin/75 μg fusion protein in 10 mM Tris–HCl, pH 7.4, containing 200 mM NaCl, 1 mM EGTA, 1 mM dithiothreitol, and 10 mM maltose at 4 °C for 16 h[f]	[62]
Procathepsin L	GST[a]		The fusion protein bound to glutathione–Sepharose 4B was digested with 0.5–2.0 units thrombin/mg fusion protein in 20 mM Tris–HCl, pH 8.4, containing 150 mM NaCl and 2.5 mM $CaCl_2$ at 20 °C for 20 h[k]	[63]
ORF-14	GST[a]		Thrombin was used to digest the fusion protein bound to glutathione–Sepharose 4B[l]	[64]
POMC	GST[a]		Thrombin (425 units/liter culture) was used to digest the fusion protein bound to Glutathione–Sepharose 4B as a slurry in 50 mM Tris–HCl, pH 8.5, containing 150 mM NaCl and 2.5 mM $CaCl_2$ at 4 °C for 1 h[l]	[65]
COOH-terminal domain of apolipoprotein-E	His-tag[b]	LVPRG	Thrombin/fusion protein, 1/100 (w/w) in 20 mM Tris–HCl, pH 7.9, containing 0.5 M imidazole and 0.5 M NaCl for 2 h at room temperature[m]	[66]
ACK-42	GST[a]		Conditions not provided	[67]
Deoxyribonucleoside kinase	GST[a]		Thrombin was used to cleave the fusion protein bound to glutathione–Sepharose 4B. Thrombin (400 units in four column volumes of 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, and 1.8 mM KH_2PO_4, pH 7.3) was recirculated through the column	[68]
Mce1A and Mce1E proteins	GST[a]		For Mce1A, 150 units of thrombin was used to cleave the fusion protein in the column eluate from the glutathione–Sepharose 4B. The digestion was performed in phosphate-buffered saline at 25 °C, time not provided. For the Mce1E fusion protein, digestion was accomplished with the fusion protein bound to the glutathione–Sepharose 4B matrix	[69]

Table 2 (*continued*)

Protein of interest	Fusion protein	Cleavage sequence	Cleavage condition	Reference
Prion domain of yeast Ure2P	GST[a]		According to manufacturer's instructions. The amount of thrombin was increased by fourfold to compensate for reduced activity in 20% glycerol	[70]
Nucleotide-binding domains of human multidrug resistance-associated protein (MRP1)	GST[a]		Cleavage with 20 units thrombin applied to the glutathione–Sepharose 4B in 20–50 mM of Tris–HCl, pH 7.5, containing 50–100 mM NaCl, 4–10 mM dithiothreitol, and 10% glycerol[n]	[71]
Nucleotide-binding domains of human multidrug resistance-associated protein (MRP1)	MBP[j]		Linker can be cleaved by either thrombin or factor Xa	[71]
Human ribosomal protein S3a	GST[a]		20 mM Hepes, pH 8.0, 150 mM NaCl, 2 mM $CaCl_2$ in ice, and 10 μg thrombin for 5 h	[72]

[a] GST, glutathione-*S*-transferase.
[b] His-tag is $His_{(6)}$.
[c] This is a relatively long cleavage sequence derived from the thrombin platelet receptor sequence and the anion-binding exosite-binding domain [73]. The thrombin cleavage site in this sequence is at the Arg-8/Ser-9 bond.
[d] The fusion protein containing the β_2-adrenoreceptor-$G_{s\alpha}$ protein with or without the thrombin cleavage sequence was obtained as a membrane preparation.
[e] Thrombin was obtained from Haematologic Technologies (Essex Junction, VT).
[f] Thrombin was obtained from Sigma Chemical (St. Louis, MO).
[g] This study described the development of an interesting assay system for the Raf/MEK/ERK kinase cascade.
[h] The thrombin used in this study was from a crude commercial source.
[i] The thrombin used in this study was obtained from CalBiochem (San Diego, CA).
[j] MBP, maltose-binding protein.
[k] Thrombin was obtained from New England Biolabs (Beverly, MA).
[l] Thrombin was obtained from Amersham–Pharmacia Biotech (Piscataway, NJ).
[m] The use of room temperature to describe reaction conditions should be avoided. It is assumed that room temperature is 23 °C.
[n] It was not possible to isolate the nonfused version of the NPD2 protein from the GST fusion protein due to proteolytic degradation.

has been a problem in the past [25,26]. More advanced purification technology [27] tends to make this less of a problem today.

This discussion would not be complete without a consideration of other approaches which have been advanced for the cleavage of fusion proteins. Tobacco etch virus protease (TEV) is a highly stringent protease [28–32]. A vector has been recently developed for use in high-throughput methods for expressing proteins for crystallography [33]. Eisenmesser et al. [34] used TEV to cleave a fusion protein between human interleukin 13 and maltose binding protein (MBP-hIL-13). Purification of the fusion protein was accomplished on amylose–agarose. After cleavage with TEV (50 mM Tris, pH 7.6, containing 200 mM NaCl, 1 mM EDTA, and 2 mM reduced glutathione), the IL-13 was insoluble and resisted further purification. Coexpression of MBP-hIL-13 and TEV resulted in high levels of hIL-13 production. Chelius et al. [35] have used TEV to obtain rat insulin-like growth factor binding protein from a thioredoxin fusion protein containing a hexahistidine sequence. Mohanty et al. [36] have examined the effect of detergents on the cleavage of fusion proteins byTEV. Some but not all zwitterionic and neutral detergents inhibited the cleavage of fusion proteins byTEV.

Approaches that involve the use of self-cleaving technology have also been advanced. Hung and Chiou [37] studied the expression of a serine protease (Tm-5) from *Trimeresurus mucrosquamatus* (Taiwan habu) in *E. coli*. A cleavage sequence for this serine protease was inserted between the poly-His coding sequence and the 5′-end of the Tm-5 clone. Renaturation of the protein after expression demonstrated that the protease had refolded correctly and cleaved the sequence between the His tail and the protease.

Hall and Kunkel [38] prepared a triple domain fusion protein represented by MHL1-VMA1-CBD, where MHL1 is the MutL homolog protein of interest, VMA1 is a splicing element (or intein), and CBD is the chitin binding domain from *Bacillus circulans*. After expression in a yeast system (*Saccharomyces cerevisiae*), a soluble cellular extract was applied to a chitin column which bound the construct through the chitin binding domain (CBD). After washing the column to remove unwanted material, self-cleavage of intein from the MHL-1 protein was induced by the addition of dithiothreitol.

Table 3
Factor Xa-catalyzed cleavage of fusion proteins

Protein of interest	Fusion partner	Cleavage sequence	Cleavage conditions	Reference
ADP-ribosylation factor	MBP[a]		20 mM Tris, pH 8.0, 1 mM EDTA, 1 mM dithiothreitol, 5 mM $MgCl_2$, and 1 mM benzamidine; 5% factor Xa (weight/weight) of fusion protein, 24–48 h at 4 °C	[52]
Tetanus toxin heavy chain	MBP[a]	IEGRISE	0.5–1.0:100 (factor Xa:protein,w/w), 10 mM Tris–HCl, pH 7.4, 200 mM NaCl, 1 mM ethylene glycol-bis(β-aminoethyl ether) N,N,N',N'-tetraacetic acid, and 1 mM dithiothreitol containing 10 mM maltose, 22 °C for 24 h. Reaction was terminated by phenylmethylsulfonyl fluoride[b]	[62]
Archaeoglobus fulgidus D-lactate dehydrogenase	MBP[a]		1 μg factor Xa/100 μg fusion protein in 20 mM Tris, 100 mM NaCl, pH 8.0, with 2 mM $CaCl_2$ for 5 days at 4 °C[c]	[74]
Soybean profilin	MBP[a]		1:20 (w/w) factor Xa/fusion protein in 100 mM phosphate-buffered saline, pH 7.6, 48 h at room temperature[d,e]	[75]
–[f]	–[f]	IEGR	10 mU factor Xa/μg fusion protein in Tris-buffered saline (10 mM Tris, pH 8.0, containing 150 mM NaCl, and TBS), 20 °C, variable times	[76]
D-Glutamate-adding enzymes	MBP[a]		Conditions recommended by New England Biolabs, 24–72 h at 4 °C[c]	[77]
Single-chain fibrin-specific antibody[g]	Hirudin		15 μg factor Xa/150 μg fusion protein, 50 mM Tris–100 mM NaCl–1 mM $CaCl_2$, pH 8.0, at room temperature[d,h]	[78]
Streptokinase	His-tag[i]	IQGR	Immobilized factor Xa in 50 mM Tris–100 mM NaCl, pH 8.0, containing 10 mM EDTA[j]	[79]
FKBP and Spo0F	His-tag[i]	IEGR	1/20 (w/w) factor Xa/fusion protein in 50 mM Tris, 100 mM NaCl, and 5 mM $CaCl_2$, pH 8.0, for 4 h at room temperature[d,k]	[80]
$p42^{IP4}$	Glutathione-*S*-transferase		1 unit factor Xa/0.1 mg fusion protein in 50 mM Tris, pH 8.0, containing 100 mM NaCl, 1 mM $CaCl_2$, and 0.5 lubrol-PX, 16 h at room temperature[c,l]	[81]
Plasminogen N-terminal peptide	His-tag[i]	IEGR	Cleavage with factor Xa was unsuccessful because of a putative second factor Xa cleavage site in the fusion protein	[82]
Human glandular kallikrein	HK Propeptide[n]	IEGR	Various factor Xa/prokallikrein ratios in 25 mM Tris, 50 mM NaCl, and 2 mM $CaCl_2$, pH 7.5, with or without 0.1% bovine serum albumin (BSA)[m]	[83]
Prohevin	MBP[a]		2–5% (w/w) factor Xa in 0.02% SDS for 72 h at 22 °C	[84]
E. coli enzyme I N-terminal domain	MBP[a]	IEGR	[c]	[85]
Staphylokinase	His-tag[i]	IEGR	60 μL factor Xa solution in 20 mM Tris, pH 8.0, containing 100 mM NaCl, 2 mM $CaCl_2$ for 9 h at 37 °C[h]	[86]
Nitrogen assimilation control protein	MBP[a]	IEGR	1 μg factor Xa/0.5 mg fusion protein in 100 mM NaH_2PO_4 (pH 7.0) with 250 mM NaCl, 2.5 mM $MgCl_2$, and 1 mM of 2-mercaptoethanol	[20]
Hepatitis B virus e antigens	Glutathione-*S*-transferase		[h][87]	
Procathepsin L	MBP[a]		Factor Xa/fusion protein (1/400, w/w) in 20 mM Tris, pH 7.5, containing 0.5 M NaCl, 10 mM maltose, and 2 mM $CaCl_2$ for 24 h at 25 °C[c]	[63]
Tyrosine hydroxylase type I	MBP[a]	IEGR	Factor Xa/fusion protein, 1/100 (w/w) in 10 mM Tris, pH 8.5, containing 200 mM NaCl and 2 mM $CaCl_2$[c]	[88]
Toxoplasma gondii ROP2 fragment	MBP[a]		2% (w/w) factor Xa in 20 mM Tris, pH 7.5, containing 200 mM NaCl, 1 mM EDTA, and 10 mM maltose[c,o]	[89]
Plasmodium falciparum acidic–basic repeat antigen	His-tag[i]		Factor Xa/fusion protein, 1/50 (w/w) in 10 mM Tris, pH 8.0, with 100 mM NaCl and 2 mM $CaCl_2$ at room temperature for 72 h[c,d]	[90]
Several proteins from *E. coli*[p]	Various fusion proteins with his-tag[i]	IEGR	1 μg factor Xa/50 μg fusion proteins in 20 mM Tris–HCl, 100 mM NaCl, and 2 mM $CaCl_2$, pH 8.0, at 23 °C for 3 h[c]	[91]
Ribosomal protein L5	MBP[a]		1 μg factor Xa/20 μg fusion protein in 10 mM Tris, pH 8.0, containing 200 mM NaCl, 5% glycerol, 1 mM EDTA, and 1 mM dithiothreitol at 4 °C for 19 h	[92]
Na^+/glucose cotransporter(SGLTI)	Glutathione-*S*-transferase		30 units factor Xa in 50 mM Tris, pH 8.0 containing 1% Triton X-100, 10 mM glutathione, and 1 mM $CaCl_2$ at 22 °C for 16 h	[93]
FKBP and Spo0F[q]	His-tag[i]	IEGR	1/20, w/w, factor Xa/fusion protein in 50 mM Tris–HCl, pH 8.0, containing 100 mM NaCl and 5 mM $CaCl_2$ at room temperature for 4 h[d,k]	[94]

Table 3 (*continued*)

Protein of interest	Fusion partner	Cleavage sequence	Cleavage conditions	Reference
Ø29 replication protein P1	MBP[a]	IEGR	1 μg factor Xa/25-100 μg fusion protein in 20 mM Tris–HCl, pH 7.5, containing 200 mM NaCl and 10 mM β-mercaptoethanol, usually at 4 °C, time not given, cleavage monitored by SDS–PAGE	[95]
Polyoma virus capsid protein VP1	His-tag[i]	IEGR	The cleavage of the fusion protein is accomplished with the fusion protein bound to the nickel affinity matrix (nickel–nitriloacetate–agarose). One hundred and eighty micrograms of factor Xa is added to the agarose matrix suspended in 20 mM Tris–HCl, 200 mM NaCl, 5% (v/v) glycerol, 2 mM EGTA, and 10 mM of 2-mercaptoethanol, pH 8.0[c]	[96]
Hypodermin C	Glutathione-*S*-transferase	IEGR	Factor Xa/fusion protein, 1/25	[97]
C1q	MBP[a]	IEGR	1 unit factor Xa/200 μg fusion protein in 20 mM Tris–HCl, pH 8.0, containing 0.1 M NaCl, 2 mM $CaCl_2$, and 5% (v/v) glycerol for 16 h (overnight) at 4 °C	[98]
Parabutoxin 3	MBP[a]	IEGR	0.5 unit factor Xa/μg fusion protein in 20 mM Tris–HCl, 100 mM NaCl, and 2 mM $CaCl_2$, pH 8.0, for 72 h at room temperature[d]	[99]
Human β-endorphin	Outer membrane protein F	IEGR	1.2 μg factor Xa/mL in 20 mM Tris–HCl, pH 7.0, at 23 °C for 12 h	[100]
Human tissue factor	MBP[a]		1:100, factor Xa:fusion protein in 20 mM Tris–HCl, 200 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol, pH 8.0, at 25 °C for 16 h	[101]

[a] MBP, maltose binding protein.
[b] Factor Xa was obtained from from Sigma Chemical (St. Louis, MO).
[c] Factor Xa was obtained from New England Biolabs (Beverly, MA).
[d] The term 'room temperature' is to be avoided in publications as such term is not definitive. At the present time, room temperature is considered to be in the range of 15–30 °C; 23 °C is a currently accepted value for the purpose of compliance with Good Laboratory Practice (GLP) requirements.
[e] Factor Xa was obtained from MobiTech (Gottingen, Germany).
[f] This work was focused on the development of some novel recombinant factor Xa preparations for use in the cleavage of fusion proteins. Two fusion proteins were used in this study: a biologically inactive α-factor-hirudin fusion protein and a maltose binding protein-calmodulin.
[g] This work describes the construction of a novel biotherapeutic which would release hirudin at the site of fibrin clot formation in vivo via cleavage of the fusion protein by factor Xa. It is suggested that this would provide "an effective localized antithrombin activity... without systemic bleeding complications."
[h] Factor Xa was obtained from Boehringer–Mannheim (Germany).
[i] His-tag is $His_{(6)}$.
[j] Immobilized factor Xa was obtained from Protein Engineering Technology (Aarhus, Denmark).
[k] Factor Xa was obtained from Promega (Madison, WI).
[l] Factor Xa was obtained from Amersham–Pharmacia Biotech (Piscataway, NJ).
[m] Recombinant factor Xa was obtained from Professor J. Stenflo (Malmo, Sweden).
[n] In this study, a factor Xa (or enterokinase) cleavage site is introduced into the propeptide region of the human glandular kallikrein precursor, activation of the recombinant protein during purification is avoided.
[o] Secondary cleavages in the fusion protein were observed in this study.
[p] This paper describes a technique study evaluating various fusion proteins with three *E. coli* proteins, NusA, GrpE, and bacterioferritin (BFR) as well as thioredoxin as fusion partners.
[q] This is a technical paper describing heterogeneity in fusion proteins reflecting partial removal of the initiating methionine residue.

Enterokinase has also been used to cleave fusion proteins. Examples include the fusion protein between *Giardia duodenalis* β-tubulin and a polyhistidine peptide [39]. *Providencia stuartii* aminoglycoside 2′-*N*-acetyltransferase and hexahistidine [40], rat mast cell protease 4 and hexahistidine [41], and human secretory leucocyte protease inhibitor and hexahistidine [42]. There are several more general references on the use of enterokinase [43,44] including a report on the expression of the enterokinase catalytic subunit in *E. coli* [44]. Several of the reports on the use of enterokinase for the cleavage of fusion proteins describe the use of green fluorescent protein inserted between the hexahistidine and the enterokinase cleavage site [44–46]. This enables the facile visualization of the fusion protein during purification.

Finally, there are situations where cyanogen bromide (CNBr) can be used for cleavage of the linker peptide [47–51]. The protein/peptide must be stable to the rather harsh solvent conditions used for CNBr cleavage.

Comment

The expression of recombinant proteins as fusion proteins where the fusion partner facilitates purification is a very useful technology. However, it is clear that rigorous characterization of the protein of interest following cleavage is essential and this is rarely done. Retention of biological activity does not necessarily mean that unwanted proteolysis has not occurred during the

cleavage step. Other issues, which need to be considered, are the choice of the solvent for the cleavage reaction with careful consideration of both pH and ionic strength. Each of the parameters can affect the rate and specificity of hydrolysis. In general, reactions involving thrombin and factor Xa are performed with near physiological conditions, with the pH in the range of 7.0–8.0 and ionic strength equivalent to that of 0.1–0.15 M NaCl. Due to its metal ion-dependent conformation, $CaCl_2$ should be included in factor Xa mediated cleavage reactions. These reaction conditions along with the ratio of enzyme to substrate, the temperature at which the reaction is performed as well as the duration of the reaction are parameters that will affect the rate and specificity of cleavage as well as the stability of the target protein. The fact that reaction conditions may require tailoring for each specific construct cannot be overemphasized. Finally, the purity of the enzyme used for cleavage is critical and, as such, the source of cleavage enzyme is critical.

References

[1] A. Hayhurst, Improved expression characteristics of single-chain Fv fragments when fused downstream of the *Escherichia coli* maltose-binding protein or upstream of a single immunoglobulin-constant domain, Protein Expr. Purif. 18 (2000) 1–10.

[2] P. Chames, J. Fieschi, D. Baty, Production of a soluble and active MBP-scFv fusion: favorable effect of the leaky tolR strain, FEBS Lett. 405 (1997) 224–228.

[3] A. Hayhurst, W.J. Harris, *Escherichia coli* skp chaperone coexpression improves solubility and page display of single-chain antibody fragments, Protein Expr. Purif. 15 (1999) 336–343.

[4] J. Schultz, Y. Lin, J. Sanderson, Y. Zuo, D. Stone, R. Mallett, S. Wilbert, D. Axworthy, A tetravalent single-chain antibody–streptavidin fusion protein for pretargeted lymphoma therapy, Cancer Res. 60 (2000) 6663–6669.

[5] B. Wick, B. Groner, Evaluation of cell surface antigens as potential targets for recombinant tumor toxins, Cancer Lett. 118 (1997) 161–172.

[6] S.H. Kim, Expression and purification of recombinant immunotoxin—a fusion protein stabilizes a single-chain Fv (scFv) in denaturing condition, Protein Expr. Purif. 27 (2003) 85–89.

[7] S.K. Sharma, D.B. Evans, A.F. Vosters, T.J. McQuade, W.G. Tarpley, Metal affinity chromatography of recombinant HIV-1 reverse transcriptase containing a human renin cleavable metal binding domain, Biotechnol. Appl. Biochem. 14 (1991) 69–81.

[8] C.F. Ford, I. Suominen, C.E. Glatz, Fusion tails for the recovery and purification of recombinant proteins, Protein Expr. Purif. 2 (1991) 95–107.

[9] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusion proteins with glutathione-*S*-transferase, Gene 67 (1988) 31–40.

[10] C. di Guan, P. Li, P.D. Riggs, H. Inouye, Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein, Gene 67 (1988) 21–30.

[11] A.P. Araujo, G. Oliva, F. Henrique-Silva, R.C. Garrett, O. Caceres, L.M. Beltramini, Influence of the histidine tail on the structure and activity of recombinant chlorocatechol 1,2-dioxygenase, Biochem. Biophys. Res. Commun. 272 (2000) 480–484.

[12] K.G. Mann, R.L. Lundblad, Biochemistry of thrombin, in: R.W. Colman, J. Hirsh, V.J. Marder, E.W. Salzman (Eds.), Hemostasis and Thrombosis: Basic Principles and Clinical Practice, Lipppincott, Philadelphia, PA, 1982, pp. 110–126.

[13] T. Takagi, R.F. Doolittle, Amino acid sequence studies on factor XIII and the peptide released during its activation by thrombin, Biochemistry 13 (1974) 750–756.

[14] K.G. Mann, Prothrombin and thrombin, in: R.W. Colman, J. Hirsh, V.J. Marder, E.W. Salzman, J.B. Lippincott (Eds.), Hemostasis and Thrombosis: Basic Principles and Clinical Practice, third ed., Lippincott, Philadelphia, PA, 1994, pp. 184–199.

[15] J.D. Lonsdale-Eccles, D.H. Hogg, D.T. Elmore, The hydrolysis of chemically modified proteins by factor Xa and thrombin, Biochim. Biophys. Acta 612 (1980) 401–409.

[16] J.Y. Chang, Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate, Eur. J. Biochem. 151 (1985) 217–224.

[17] J.Y. Chang, S.S. Alkan, N. Hilschmann, D.G. Braun, Thrombin specificity. Selective cleavage of antibody light chains at the joints of variable with joining regions and joining with constant regions, Eur. J. Biochem. 151 (1985) 225–230.

[18] M.J. Raftery, L. Collinson, C.L. Geczy, Overexpression, oxidative refolding, and zinc binding of recombinant forms of murine S100 protein MRP14(S100A9), Protein Expr. Purif. 15 (1999) 228–235.

[19] J. Zaitseva, H. Zhang, B.A. Binnie, M. Hermodson, The proteins encoded by the rbs operon of *Escherichia coli*: II. Use of chimeric protein constructs to isolate and characterize RbsC, Protein Sci. 5 (1996) 1100–1107.

[20] W.B. Muse, R.A. Bender, The amino-terminal 100 residues of the nitrogen assimilation control protein (NAC) encode all known properties of NAC from *Klebsiella aerogenes* and *Escherichia coli*, J. Bacteriol. 181 (1999) 934–940.

[21] A. Belmouden, K.H.D. Le, F. Lederer, H.-J. Garchon, Molecular cloning and nucleotide sequence of cDNA encoding rat kidney long-chain L-2-hydroxy acid oxidase. Expression of the catalytically active recombinant protein as a chimera, Eur. J. Biochem. 214 (1993) 17–25.

[22] Y.H. Ko, P.J. Thomas, M.R. Delannoy, P.L. Pedersen, The cystic fibrosis transmembrane conductance regulator. Overexpression, purification, and characterization of wild type and ΔzF508 mutant forms of the first nucleotide binding fold in fusion with the maltose-binding protein, J. Biol. Chem. 268 (1993) 24330–24338.

[23] S. Sharma, D.R. Rose, Cloning, overexpression, purification, and characterization of the carboxyl-terminal nucleotide binding domain of *P*-glycoprotein, J. Biol. Chem. 270 (1995) 14085–14093.

[24] C. Wang, A.F. Castro, D.M. Wilkes, G.A. Altenberg, Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: comparison of fusions to glutathione *S*-transferase, thioredoxin and maltose-binding protein, Biochem. J. 338 (1999) 77–81.

[25] S. Suzuki, N. Sakuragawa, A study on the properties of commercial thrombin preparations, Thromb. Res. 53 (1989) 271–277.

[26] V.H. Donaldson, J. Kleniewski, The role of plasmin in kinin-release by preparations of thrombin, Thromb. Res. 16 (1979) 401–406.

[27] R.D. Litwiller, R.J. Jenny, K.G. Mann, Identification and isolation of vitamin K-dependent proteins by HPLC, Anal. Biochem. 158 (1986) 355–368.

[28] J. Phan, A. Zanov, A.G. Evdokimov, J.E. Tropea, H.K. Peters 3rd, R.B. Kapust, M. Li, A. Wlodawer, D.S. Waugh, Structural basis for the substrate specificity of tobacco etch virus protease, J. Biol. Chem. 277 (2002) 50564–50572.

[29] R.G. Kapust, J. Tozser, T.D. Copeland, D.S. Waugh, The P1′ specificity of tobacco etch virus protease, Biochem. Biophys. Res. Commun. 294 (2002) 949–955.
[30] K.N. Faber, A.M. Kram, M. Ehrmann, M.A. Veenhuis, A novel method to determine the topology of peroxisomal membrane proteins in vivo using the tobacco etch virus protease, J. Biol. Chem. 276 (2001) 36501–36507.
[31] L.J. Lucast, R.T. Batey, J.A. Doudna, Large-scale purification of a stable form of recombinant tobacco etch virus protease, Biotechniques 30 (2001) 544–546.
[32] J. Haspel, C. Blanco, J. Jacob, M. Grumet, System for cleavable Fc fusion proteins using tobacco etch virus (TEV) protease, Biotechniques 30 (2001) 60–61.
[33] L. Stols, M. Gu, L. Dieckman, R. Raffen, F.R. Collart, M.I. Donnelly, A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site, Protein Expr. Purif. 25 (2002) 8–15.
[34] E.Z. Eisenmesser, R.B. Kapust, J.P. Nawrocki, M.J. Mazzulla, M.L.K. Pannell, D.S. Waugh, R.A. Byrd, Expression, purification, refolding and characterization of recombinant human interleukin-13: utilization of intracellular processing, Protein Expr. Purif. 20 (2000) 186–195.
[35] D. Chelius, M.A. Baldwin, X. Lu, X.E.M. Spencer, Expression, purification and characterization of the structure and disulfide linkages of insulin-like growth factor binding protein-4, J. Endocrinol. 168 (2001) 283–296.
[36] A.K. Mohanty, C.R. Simmons, M.C. Werner, Inhibition of tobacco etch virus protease activity by detergents, Protein Expr. Purif. 27 (2003) 109–114.
[37] C.C. Hung, S.H. Chiou, Expression of a kallikrein-like protease from the snake venom: engineering of autocatalytic site in the fusion protein to facilitate protein refolding, Biochem. Biophys. Res. Commun. 275 (2000) 924–930.
[38] M.C. Hall, T.A. Kunkel, Purification of eucaryotic MutL homologs from *Saccharomyces cerevisiae* using self-cleaving affinity technology, Protein Expr. Purif. 21 (2001) 333–342.
[39] L.M. MacDonald, A. Armson, R.C. Thompson, J.A. Reynoldson, Expression of *Giardia duodenalis* beta-tubulin as a soluble protein in *Escherichia coli*, Protein Expr. Purif. 22 (2001) 25–30.
[40] K. Franklin, A.J. Clarke, Overexpression and characterization of the chromosomal aminoglycoside 2′-*N*-acetyltransferase of *Providencia stuartii*, Antimicrob. Agents Chemother. 45 (2001) 2238–2244.
[41] J. Karlson, G. Pejler, G. Froman, L. Hellman, Rat mast cell protease 4 is a beta-chymase with unusually stringent substrate recognition profile, J. Biol. Chem. 277 (2002) 18579–18585.
[42] L.R. Gray, A.L. Alexander, D.C. Shugars, Construction, nondenaturing affinity purification, and characterization of baculovirally expressed human secretory leukocyte protease inhibitor, Protein Expr. Purif. 26 (2002) 179–186.
[43] S.I. Choi, H.W. Song, J.W. Moon, B.L. Seong, Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology, Biotechnol. Bioeng. 75 (2001) 718–724.
[44] L.D. Yuan, Z.C. Hua, Expression, purification, and characterization of a biologically active bovine enterokinase catalytic subunit in *Escherichia coli*, Protein Expr. Purif. 25 (2002) 300–304.
[45] C.F. Wu, H.J. Cha, J.J. Valdes, W.E. Bentley, GFP-visualized immobilized enzymes: degradation of paraoxin via organophosphorous hydrolase in a packed column, Biotechnol. Bioeng. 77 (2002) 212–218.
[46] J. Lu, H. Xu, W.K. Herber, W.E. Bentley, G. Rao, Integrated bioprocessing in *Saccharomyces cerevisiae* using green fluorescent protein as a fusion partner, Biotechnol. Bioeng. 79 (2002) 682–693.
[47] A. Kuliopulos, C.T. Walsh, Production, purification, and cleavage of tandem repeats of recombinant peptides, J. Am. Chem. Soc. 116 (1994) 4599–4607.
[48] H. Döbeli, H. Andres, N. Breyer, N. Draeger, D. Sizmann, M.T. Zuber, M.B. Weinert, B. Wipf, Recombinant fusion proteins for the industrial production of disulfide bridge containing peptides: purification, oxidation without concatamer formation, and selective cleavage, Protein Expr. Purif. 12 (1998) 404–414.
[49] F.M. Assadi-Porter, D.J. Aceti, H. Cheng, J.L. Markley, Efficient production of recombinant brazzein, a small, heat-stable, sweet-tasting protein of plant origin, Arch. Biochem. Biophys. 376 (2000) 252–258.
[50] F. Karamloo, S. Scheurer, A. Wangosch, S. May, D. Haustein, S. Vieths, Pyr c 1, the major allergen from pear (*Pyrus communia*), is a new member of the Bet v 1 allergen family, J. Chromatogr. B Biomed. Sci. Appl. 756 (2001) 281–293.
[51] W.D. Fairlie, A.D. Uboldi, D.P. De Souza, G.J. Hemmings, N.A. Nicola, M. Baca, A fusion protein system for the recombinant production of short disulfide-containing peptides, Protein Expr. Purif. 26 (2002) 171–178.
[52] J. Moss, R.S. Haun, S.-C. Tsai, S.C.F. Welsh, F.-J.S. Lee, S.R. Price, M. Vaughan, Activation of cholera toxin by ADP-ribosylation factors: 20 kDa nucleotide binding proteins, Methods Enzymol. 237 (1994) 44–63.
[53] R. Seifert, K. Wenzel-Seifert, U. Gether, V.T. Lam, B.K. Kobilka, Examining the efficiency of receptor/G-protein coupling with a cleavable β_2-adrenoreceptor-$G_{s\alpha}$ fusion protein, Eur. J. Biochem. 260 (1999) 661–666.
[54] G.J.P. Fernando, B. Murray, J. Zhou, I.H. Frazier, Expression, purification and immunological characterization of the transforming protein E7, from cervical cancer-associated human papillomavirus type 16, Clin. Exp. Immunol. 115 (1999) 397–403.
[55] O.B. McDonald, W.J. Chen, B. Ellis, C. Hoffman, L. Overton, M. Rink, A. Smith, C.J. Marshall, E.R. Wood, A scintillation proximity assay for the Raf/MEK/ERK kinase cascade: high-throughput screening and identification of selective enzyme inhibitors, Anal. Biochem. 268 (1999) 318–329.
[56] H. Hotta, S. Kon, Y.U. Katagiri, N. Tosa, T. Tsukamoto, A.F. Chambers, T. Uede, Detection of various epitopes of murine osteopontin by monoclonal antibodies, Biochem. Biophys. Res. Commun. 257 (1999) 6–11.
[57] E. Schleiff, R. Khanna, S. Orlicky, A. Vrielink, Expression, purification, and in vitro characterization of the *HuProteinman* outer mitochondrial membrane receptor human translocase of the outer mitochondrial membrane 20, Arch. Biochem. Biophys. 367 (1999) 95–103.
[58] N.T. Lu, P.L. Pedersen, Cystic fibrosis transmembrane conductance regulator: the purified NBF1 + R protein interacts with the purified NBF2 domain to form a stable NGF1 + F/NBF2 complex while inducing a conformational change transmitted to the C-terminal region, Arch. Biochem. Biophys. 375 (2000) 7–20.
[59] J. Srinivassan, M. Schachner, W.A. Catterall, Interaction of voltage-gated sodium channels with the extracellular matrix molecules tenascin-C and tenascin-R, Proc. Natl. Acad. Sci. USA 95 (1998) 15753–15757.
[60] T. Jostock, G. Blinn, C. Renné, K.-J. Kallen, S. Rose-John, S.J. Müllberg, Immunoadhesins of interleukin-6 and the IL-6/soluble IL-6R fusion protein hyper-IL-6, J. Immunol. Methods 223 (1999) 171–183.
[61] S.K. Shenoy, L. Yu, L.C.-a. Yu, Identification of quinone-binding and heme-ligating residues of the smallest membrane-anchoring subunit(QPs3) of bovine heart mitochondrial succinate:ubiquinone reductase, J. Biol. Chem. 274 (1999) 8717–8722.
[62] Y. Li, R. Aoki, J.O. Dolly, Expression and characterization of the heavy chain of tetanus toxin: reconstitution of the fully-recombinant dichain protein in active form, J. Biochem. 125 (1999) 1200–1208.

[63] T. Ogino, T. Kaji, M. Kawabata, K. Satoh, K. Tomoo, T. Ishida, H. Yamazaki, K. Ishidoh, E. Kominami, Function of the propeptide region in recombinant expression of active procathepsin L in *Escherichia coli*, J. Biochem. 126 (1999) 78–83.

[64] S. Ahmad, H.A. Amoudy, J.E.R. Thole, D.B. Young, A. Mustafa, Identification of a novel protein antigen encoded by a *Mycobacterium tuberculosis*-specific RD1 region gene, Scand. J. Immunol. 49 (1999) 515–522.

[65] C. Bert, V.V. Vijver, M. Andries, P. Verhaert, P. Proost, B. De Vreese, J. Van Beeumen, H. Vankelecom, C. Denef, Production of recombinant rat proopiomelanocortin1–74 and characterization of its mitogenic action on pituitary lactotrophs, Mol. Cell. Endocrinol. 154 (1999) 111–122.

[66] M. Forstner, C. Peters-Libeu, E. Contreras-Forest, Y. Newhouse, M. Knapp, B. Rupp, K.H. Weisgraber, Carboxyl-terminal domain of human apoprotein E: expression, purification and crystallization, Protein Expr. Purif. 17 (1999) 267–272.

[67] M.S.A. Nur-E-Kamal, J.M. Kamal, M.M. Qureshi, H. Maruta, The CDC42-specific inhibitor derived from ACK-1 blocks v-HA-Ras-induced transformation, Oncogene 18 (1999) 7787–7793.

[68] B.M. Munch-Petersen, W. Knech, C. Lenz, L. Søndergaard, J. Piškur, Functional expression of a multisubstrate deoxyribonucleoside kinase from *Drosophila melanogaster* and its C-terminal deletion mutants, J. Biol. Chem. 275 (2000) 6673–6679.

[69] S. Ahmad, P.K. Akbar, H.G. Wiker, M. Harboe, A.S. Mustafa, Cloning, expression and immunological reactivity of two mammalian cell entry proteins encoded by the *mce1* operon of *Mycobacterium tuberculosis*, Scand. J. Immunol. 50 (1999) 510–518.

[70] M. Schlumpberger, H. Wille, M.A. Baldwin, D.A. Butler, I. Herskowitz, S.B. Pruisner, The prion domain of yeast Ure2p induces autocatalytic formation of amyloid fibers by a recombinant fusion protein, Protein Sci. 9 (2000) 440–451.

[71] R.H. Cool, M.K. Veenstra, W. van Klompenburg, R.I.R. Heyne, M. Müller, E.G.E. de Vries, H.W. van Veen, W.N. Konings, *S*-Decyl-glutathione nonspecifically stimulates the ATPase activity of the nucleotide-binding domains of the human multidrug resistance-associated protein, MRP1 (ABCC1), Eur. J. Biochem. 269 (2002) 3470–3478.

[72] D. Song, S. Sakamoto, T. Taniguchi, Inhibition of poly (ADP-ribose) polymerase activity by Bcl-2 in association with the ribosomal protein S3a, Biochemistry 41 (2002) 929–934.

[73] S.R. Couglin, T.-K.H. Vu, D.T. Hung, V.I. Wheaton, Characterization of a functional thrombin receptor. Issues and opportunities, J. Clin. Invest. 89 (1992) 351–355.

[74] D.W. Reed, P.L. Hartzell, The *Archaeoglobus fulgidus* D-lactate dehydrogenase is a Zn^{2+} flavoprotein, J. Bacteriol. 184 (1999) 7580–7587.

[75] H.P. Rihs, Z. Chen, R. Ruëff, A. Petersen, P. Rozynek, H. Heimann, X. Baur, IgE binding of the recombinant allergen soybean protein (rGly m3) is mediated by conformational epitopes, J. Allergy Clin. Immunol. 104 (1999) 1293–1301.

[76] M.M. Guarna, H.C.F. Côte, E.M. Kwan, G.L. Rintoul, B. Meyhack, J. Heim, R.T.A. MacGillivray, R.A.J. Warren, D.G. Kilburn, Factor X fusion proteins: improved production and use in the release in vitro of biologically active hirudin from an inactive α-factor-hirudin fusion protein, Protein Expr. Purif. 20 (2000) 133–141.

[77] A.W. Walsh, P.J. Falk, J. Thanassi, L. Discotto, M.J. Pucci, H.-T. Ho, Comparison of the D-glutamate-adding enzymes from selected gram-positive and gram-negative bacteria, J. Bacteriol. 181 (1999) 5395–5401.

[78] K. Peter, J. Graeber, S. Kipriyanov, M. Zewe-Welschof, M.S. Runge, W. Kübler, M. Little, C. Bode, Construction and functional evaluation of a single-chain antibody fusion protein with fibrin targeting and thrombin inhibition after activation by factor Xa, Circulation 101 (2000) 1158–1164.

[79] L.B. Johnsen, L.K. Rasmussen, T.E. Petersen, M. Etaerodt, S.N. Fedosov, Kinetic and structural characterization of a two-domain streptokinase: dissection of domain functionality, Biochemistry 39 (2000) 6440–6448.

[80] Z. Yan, G.W. Caldwell, P.A. McDonell, W.J. Jones, A. August, J.A. Masucci, Mass spectrometric determination of a novel modification of the N-terminus of histidine-tagged proteins expressed in bacteria, Biochem. Biophys. Res. Commun. 259 (1999) 217–282.

[81] T. Hanck, R. Stricker, U.M. Krishna, J.R. Falck, Y.-T. Chang, S.-K. Chung, G. Reiser, Recombinant $p42^{IP4}$, a brain-specific 42 kDa high-affinity Ing(1,3,4,5)P_4 receptor protein, specifically interacts with lipid membranes containing Ptd-Ins(3,4,5)P_3, Eur. J. Biochem. 261 (1999) 577–584.

[82] V.O. Lewis, M. Gehrmann, L. Weissbach, J.E. Hyman, A. Rielly, D.G. Jones, M. Llinás, J. Schaller, Homologous plasminogen N-terminal and plasminogen-related gene A and B peptides. Characterization of cDNAs and recombinant fusion proteins, Eur. J. Biochem. 259 (1999) 618–625.

[83] J. Lövgren, S. Tian, A. Lundwall, M. Karp, H. Lilja, Production and activation of recombinant hK2 with propeptide mutation resulting in high expression levels, Eur. J. Biochem. 266 (1999) 1050–1055.

[84] P. Rozynek, A. Posch, X. Baur, Cloning, expression and characterization of the major latex allergen prohevin, Clin. Exp. Allergy 28 (1998) 1418–1426.

[85] A. Fomenkov, A. Valiakhmetov, L. Brand, S. Roseman, In vivo and in vitro complementation of the N-terminal domain of enzyme I of the *Escherichia coli* phosphotransferase system by the cloned C-terminal domain, Proc. Natl. Acad. Sci. USA 95 (1998) 8491–8495.

[86] B. Schlott, K.-H. Gührs, M. Hartmann, A. Röcker, D. Collen, NH_2-terminal structural motifs in staphylokinase required for plasminogen activation, J. Biol. Chem. 273 (1998) 22346–22350.

[87] G.-Y. Hwang, J.-C. Wang, C.-C. Wu, Immunological characterization of two major secreted forms of recombinant hepatitis B virus e antigens, Virus Res. 59 (1999) 203–210.

[88] A. Nakashima, K. Mori, T. Nagatus, A. Ota, Expression of human tyrosine hydroxylase type I in *Escherichia coli* as a protease-cleavable fusion protein, J. Neural Transm. 106 (1999) 819–824.

[89] A. Jacquet, V. Daminet, M. Haumont, L. Garcia, S. Chaudoir, A. Bollen, R. Biemans, Expression of a recombinant *Toxoplasma gondii* ROP2 fragment as a fusion protein in bacteria circumvents insolubility and proteolytic degradation, Protein Expr. Purif. 17 (1999) 392–400.

[90] A. Kushwaha, P.P.L. Rao, V.S. Duttu, P. Malhotra, V.S. Chauhan, Expression and characterisation of *Plasmodium falciparum* acidic basic repeat antigen expressed in *Escherichia coli*, Mol. Biochem. Parasitol. 106 (2000) 213–224.

[91] G.D. Davis, C. Elisee, D.M. Newham, R.G. Harrison, New fusion protein systems designed to give soluble expression in *Escherichia coli*, Biotechnol. Bioeng. 65 (1999) 382–388.

[92] O.N. Pakhomova, L.-C.C. Yeh, J. Monette, J.C. Lee, In vitro assembly of yeast 5S rRNA and a fusion protein containing ribosomal protein L5 and maltose binding protein, Biochimie 81 (1999) 1015–1023.

[93] M. Panayotova-Heiermann, D.W. Leung, B.A. Hirayama, E.M. Wright, Purification and functional reconstitution of a truncated human Na^+/glucose cotransporter (SGLT1) expressed in *E. coli*, FEBS Lett. 359 (1999) 386–390.

[94] Z. Yan, G.W. Caldwell, P.A. McDonnel, W.J. Jones, A. August, J.A. Masucci, Mass spectrometric determination of a novel

modification of the N-terminus of histidine-tagged proteins expressed in bacteria, Biochem. Biophys. Res. Commun. 259 (1999) 271–282.

[95] A. Bravo, M. Salas, Polymerization of bacteriophage ø29 replication protein p1 into profilament sheets, EMBO J. 17 (1998) 6096–6105.

[96] H. Braun, K. Boller, J. Löwer, W.M. Bertling, A. Zimmer, Oligonucleotide and plasmid DNA packaging into polyoma VPI virus-like particles expressed in *Escherichia coli*, Biotechnol. Appl. Biochem. 29 (1999) 31–43.

[97] R. Casais, J.M.M. Alonso, J.A. Boga, F. Parra, *Hydoerma lineatum*: expression of enzymatically active hypodermin C in *Escherichia coli* and its use for the immunodiagnosis of hypodermosis, Exp. Parasitol. 90 (1998) 14–19.

[98] U. Kishore, L.E.A. Leigh, P. Eggleton, P. Strong, M.V. Perdikoulis, A.C. Willis, K.B.M. Reid, Functional characterization of a recombinant form of the C-terminal, globular head region of the B-chain of human serum complement protein, C1q, Biochem. J. 333 (1998) 27–32.

[99] I. Huys, K. Dyason, E. Waelkens, F. Verdonck, J. van Zy, J. du Plessis, G.J. Müller, J. vander Walt, E. Clynen, L. Schoofs, L.J. Tytgat, Purification, characterization and biosynthesis of parabutoxin 3, a component of *Parabuthus transvaalicus* venom, Eur. J. Biochem. 269 (2002) 1854–1865.

[100] K.J. Jeong, S.Y. Lee, Excretion of human β-endorphin into culture medium by using outer membrane protein F as a fusion partner in recombinant *Escherichia coli*, Appl. Environ. Microbiol. 68 (2002) 4979–4985.

[101] M. Guan, B. Su, C. Ye, Y. Lu, Production of extracellular domain of human tissue factor using maltose-binding protein fusion system, Protein Expr. Purif. 26 (2002) 229–234.

US005298599A

United States Patent [19]

Rezaie et al.

[11] Patent Number: 5,298,599

[45] Date of Patent: Mar. 29, 1994

[54] **EXPRESSION AND PURIFICATION OF RECOMBINANT SOLUBLE TISSUE FACTOR**

[75] Inventors: **Alireza Rezaie**, Moore; **Charles T. Esmon**; **James H. Morrissey**, both of Oklahoma City, all of Okla.

[73] Assignee: **Oklahoma Medical Research Foundation**, Oklahoma City, Okla.

[21] Appl. No.: **816,679**

[22] Filed: **Jan. 3, 1992**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 730,040, Jul. 12, 1991, Pat. No. 5,202,253, which is a continuation of Ser. No. 292,447, Dec. 30, 1988, abandoned, which is a continuation-in-part of Ser. No. 683,682, Apr. 10, 1991.

[51] **Int. Cl.**[5] **C07K 13/00**

[52] **U.S. Cl.** **530/350**; 435/68.1; 435/69.7; 435/69.6; 530/381; 530/413; 530/388.25; 530/413

[58] **Field of Search** 435/68.1, 69.7; 530/350, 413; 536/27

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,506,009	3/1985	Linhoff et al.	435/7
4,851,341	7/1989	Hopp et al.	435/68
5,147,638	9/1992	Esmon et al.	424/85.8

FOREIGN PATENT DOCUMENTS

0278776	2/1988	European Pat. Off. .
296413	12/1988	European Pat. Off. .
WO88/04692	12/1987	PCT Int'l Appl. .
WO88/07543	10/1988	PCT Int'l Appl. .
WO88/09817	12/1988	PCT Int'l Appl. .
WO90/07524	12/1989	PCT Int'l Appl. .

OTHER PUBLICATIONS

Long, G. L. (1987) J. Cell. Biochem. 33: 185–90.

Vigano-D'Angelo, et al., (1986) J. Clin. Invest. 77: 416–25.

Morrissey, J. H., et al. (1987) Cell 50: 129–35.

Ghrayeb, J., et al. "Secretion Cloning Vectors in *Escherichia coli*," EMBO Journal, 3(10), 2437–2442 (1984).

Lei, Shau-Ping, et al., "Characterization of the *Erwinia carotovora pelB* Gene and its Product Pectate lyase," J. Bacteriol., 169(9), 4379–4383 (1987).

Pickett, S. K., et al., "A Calcium-Dependent Antibody for Identification and Purification of Recombinant Proteins," Bio Techniques, 7(6) (1989).

deSousa, et al., *Thrombosis Research*, vol. 51, No. 2, pp. 165–173 (1988).

Poggio, et al., *Thrombosis and Haemostasis*, vol. 65, No. 2, pp. 160–164 (1991).

Spicer, et al., *Proceedings of the National Academy of Sciences U.S.A.*, vol. 84, pp. 5148–5152 (Aug. 1987).

Scarpati, et al., *Biochemistry*, vol. 26, No. 17, pp. 5234–5238 (1987).

Fisher, et al., *Thrombosis Research*, vol. 48, No. 1, pp. 89–99 (1987).

(List continued on next page.)

Primary Examiner—Robert J. Hill, Jr.
Assistant Examiner—David L. Fitzgerald
Attorney, Agent, or Firm—Kilpatrick & Cody

[57] **ABSTRACT**

A method is disclosed to make any protein in a form that can be isolated rapidly from a solution using a specific monoclonal antibody designated "HPC-4". It has now been determined that it is possible to form a fusion protein of the epitope with a protein to be isolated, and isolate the protein using HPC-4-based affinity chromatography. In the preferred embodiment, a specific protease cleavage site is inserted between the epitope and the protein so that the epitope can be easily removed from the isolated protein. In an example, a functionally active soluble tissue factor including the twelve amino acid epitope recognized in combination with calcium by HPC-4 and a factor Xa cleavage site was expressed from a vector inserted into a procaryotic expression system. The recombinant tissue factor can be rapidly isolated in a single chromatographic step using the HPC-4 monoclonal antibody immobilized on a suitable substrate. Once isolated, the Protein C epitope is removed by cleavage with factor Xa, leaving the functionally active, soluble tissue factor.

6 Claims, 3 Drawing Sheets

OTHER PUBLICATIONS

Paborsky, et al., *Biochemistry*, vol. 28, No. 20, pp. 8072–8077 (1989).

Paborsky and Harris, Thrombosis Research, vol. 60, No. 5, pp. 367–376 (1990).

Ruff, et al., *Thrombosis and Haemostasis*, vol. 62, p. 347 (1989).

Ruff, et al., *Journal Biological Chemistry*, vol. 266, No. 4, pp. 2158–2166 (Feb. 5, 1991).

Hoffman, et al., *Journal of Laboratory Clinical MEdicine*, vol. 111, No. 4, pp. 475–481 (Apr. 1988).

Gordon, et al., *Journal of Laboratory Clinical Medicine*, vol. 109, No. 3, pp. 409–413 (Apr. 1987).

K. A. Mitropoulos, *Seminars in Thrombosis and Hemostasis*, vol. 14, No. 3, pp. 246–252 (1988).

Scarabin, et al., *Thrombosis Research*, vol. 45, No. 6, pp. 845–850 (1987).

Esmon et al., Develop. Biol. Standard., 67: 75–82, 1987.

Dreyfus et al., New England Journal of Medecine 325: 1565–1568, 1991.

Seligsohn, et al., The New England Journal of Medicine, vol. 310 No. 9, pp. 559–562 (Mar. 1, 1984).

Vukovich, et al., *British Journal of Haematology* 70, pp. 435–440 (1988).

C. T. Esmon, et al., Joint IABS/CSL Symposium on Standarization in Blood Fractionation including Coagulation Factors, Melbourne, Australia, 1986 *Develop. biol. Standard*, vol. 67, pp. 51–57 (S. Karger, Basel, 1987).

C. T. Esmon, et al., *Journ. of Biol. Chem.* vol. 264. No. 9 pp. 4743–4746 (Mar. 25, 1989).

C. T. Esmon, *Science* vol. 235, pp. 1348–1352 (Mar. 13, 1987).

A. K. Ohlin, et al., *Journ. of Biol. Chem.* vol. 262, No. 28 pp. 13798–13804 (Oct. 5, 1987).

Laurell, et al., *FEBS Letters* vol. 191, No. 1, pp. 75–81 (Oct. 1985).

Abstract 846, Ohlin, et al., *Thrombosis and Haemostasis* 58 (1): 230 (Jul. 1987).

Abstract 848, Freyssinet, et al., *Thrombosis and Haemostasis* 58(1): 230 (Jul. 1987).

Suzuki, et al., *J. Biochem.* 97, 127–138 (1985).

Stearns, et al., *Journ of Biol. Chem.*, vol. 263, No. 2, pp. 826–832 (Jan. 15, 1988).

Beckmann, et al., *Nucl. Acids Res.* vol. 13, No. 14 pp. 5233–5247 ((1985).

Taylor et al. J. Clin Invest. 79: 918–25 1987.

Ikeda et al. Thromb. Res. 39: 297–306 1985.

Goding et al., "Monoclonal Antibodies" Academic Press 1983 pp. 111–113.

Sugo et al. Thromb. Hernost. 58(10 229 1987.

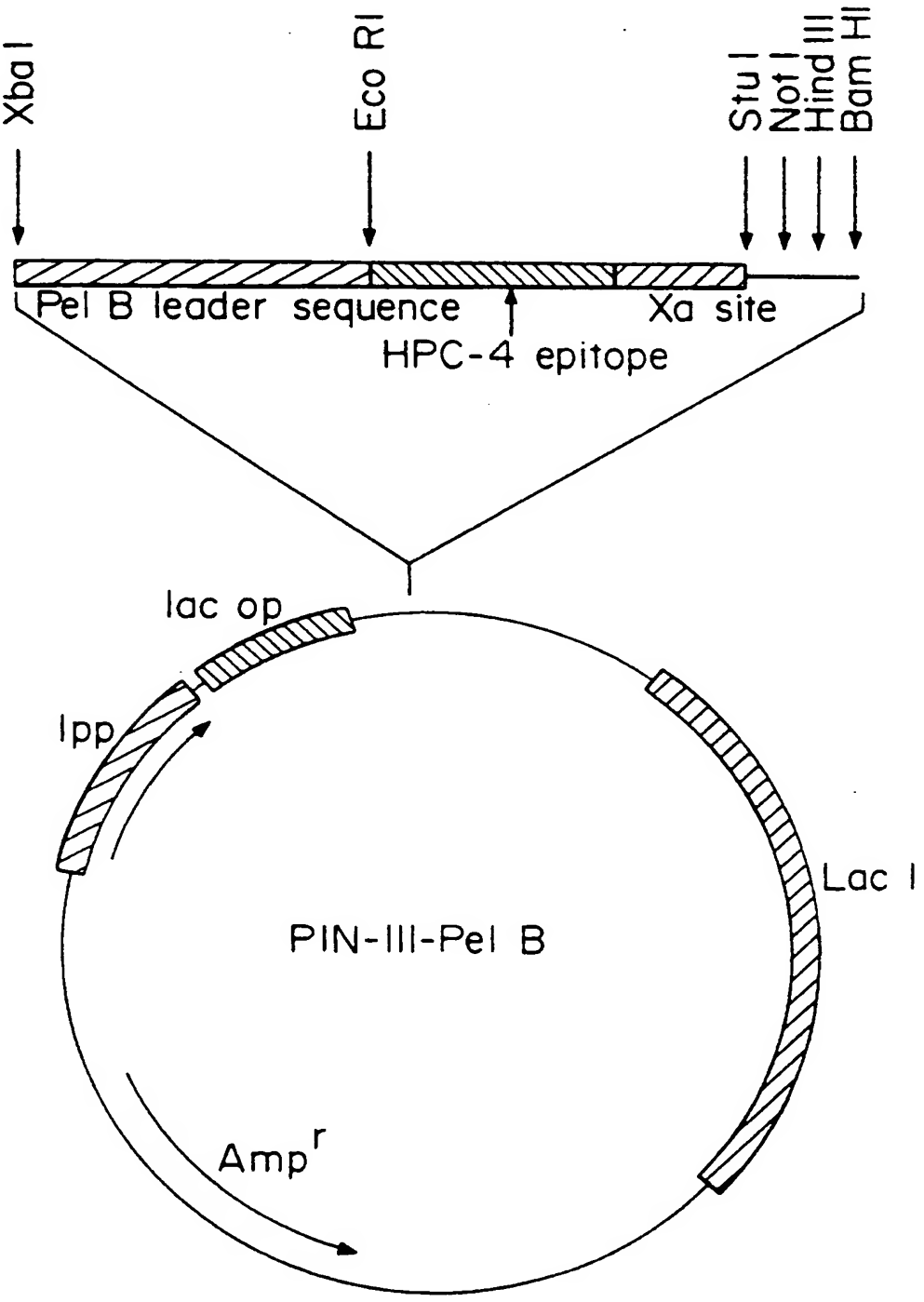

Xba I
Eco RI
Stu I
Not I
Hind III
Bam HI
Pel B leader sequence
Xa site
HPC-4 epitope
lac op
lpp
Lac I
PIN-III-Pel B
Amp^r

FIG. 1

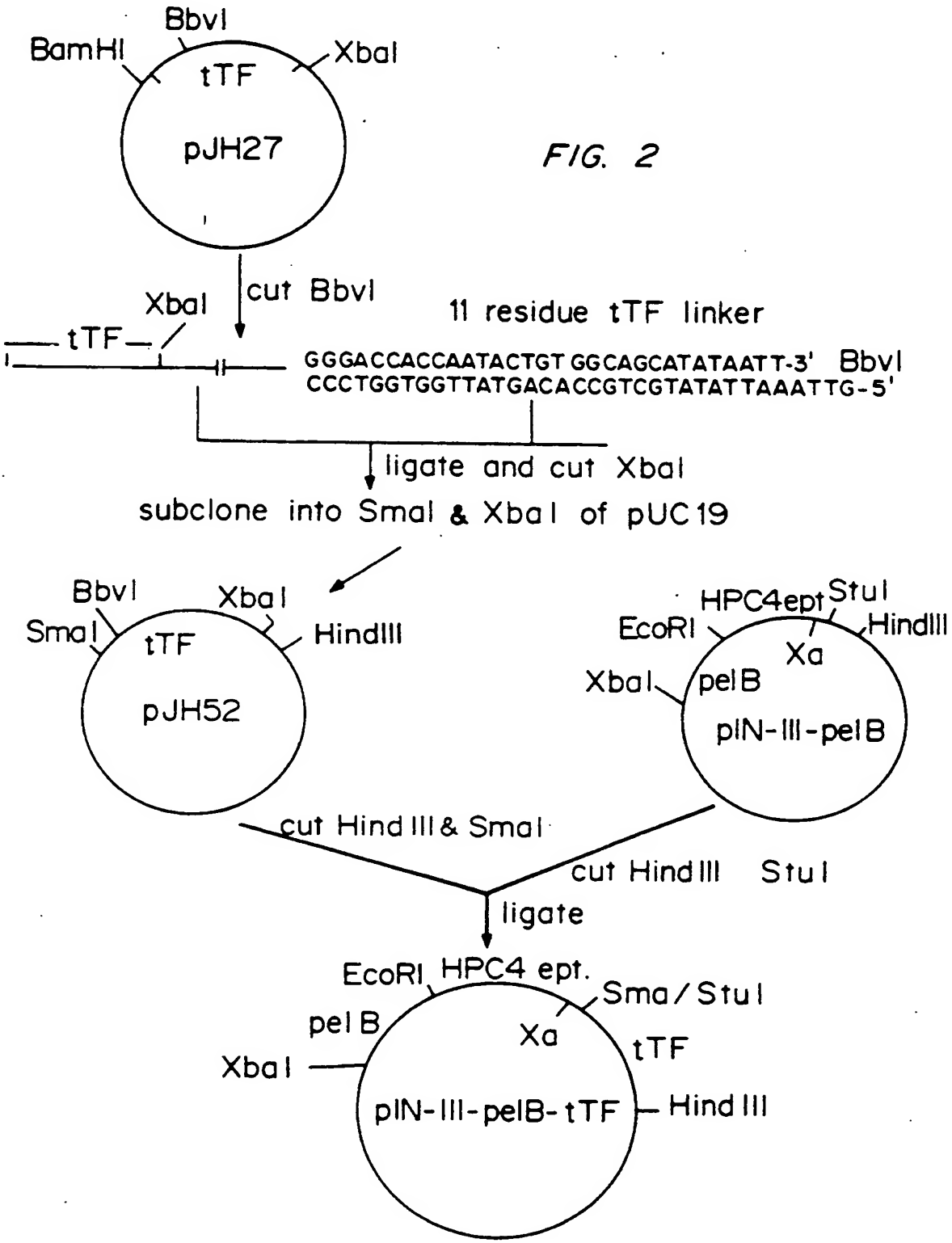

Bbvl
BamHI
Xbal
tTF
pJH27
cut Bbvl
Xbal
tTF
11 residue tTF linker
GGGACCACCAATACTGT GGCAGCATATAATT-3' Bbvl
CCCTGGTGGTTATGACACCGTCGTATATTAAATTG-5'
ligate and cut Xbal
subclone into SmaI & XbaI of pUC19
Bbvl
Xbal
SmaI
tTF
HindIII
pJH52
HPC4ept
StuI
EcoRI
HindIII
Xa
XbaI
pelB
pIN-III-pelB
cut HindIII & SmaI
cut HindIII StuI
ligate
EcoRI
HPC4 ept.
Sma/StuI
pelB
Xa
tTF
XbaI
pIN-III-pelB-tTF
HindIII

FIG. 2

FIG. 3

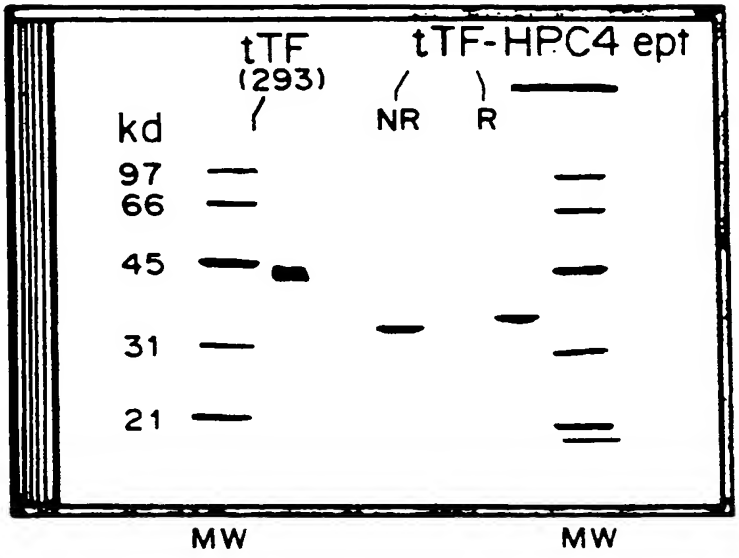
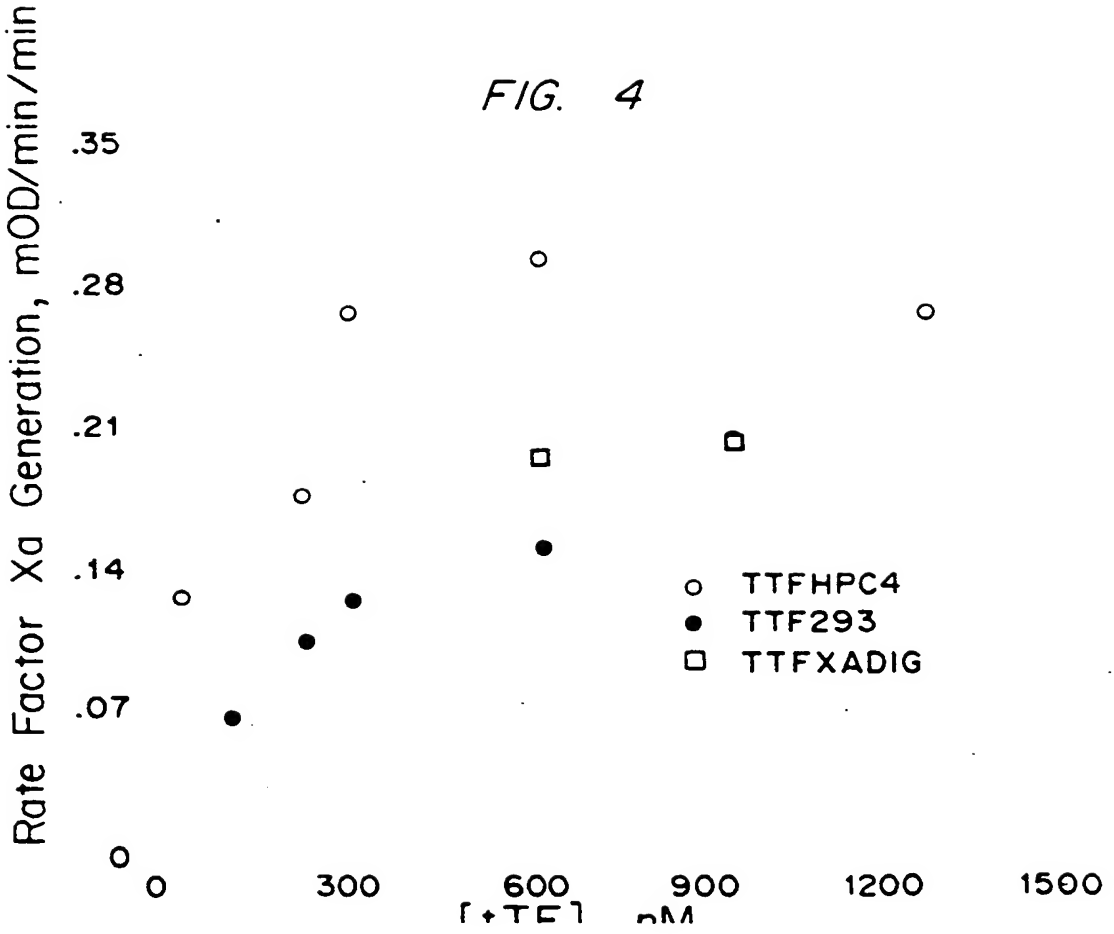

tTF
(293)
tTF-HPC4 ept
NR
R
kd
97
66
45
31
21
MW
MW

FIG. 4


Rate Factor Xa Generation, mOD/min/min
.35
.28
.21
.14
.07
0
0
300
600
900
1200
1500
TTFHPC4
TTF293
TTFXADIG

EXPRESSION AND PURIFICATION OF RECOMBINANT SOLUBLE TISSUE FACTOR

United States government has rights in this invention by virtue of grants from the National Institutes of Health, grant numbers R01 HL44225 and R01 HL29807.

This is a continuation-in-part of. U.S. Ser. No. 07/730,040 filed Jul. 12, 1991, now U.S. Pat. No. 5,202,253, which is a continuation of U.S. Ser. No. 07/292,447, abandoned, entitled "Monoclonal Antibody against Protein C" filed Dec. 30, 1988 by Charles T. Esmon and Naomi L. Esmon, and a continuation-in-part of pending U.S. Ser. No. 07/683,682 entitled "Quantitative Clotting Assay for Activated Factor VII" filed Apr. 10, 1991 by James H. Morrissey.

This invention is generally in the area of methods for purifying proteins, especially blood clotting proteins, using recombinant technology and an unique epitope of a monoclonal antibody directed against Protein C zymogen.

Methods for purifying proteins have been used for many years and can be generally divided into chromatographic methods, for example, ion exchange chromatography, molecular weight sieving, high pressure liquid chromatography, affinity chromatography, and electrophoretic methods, for example, electrophoresis on agarose or acrylamide gels and isoelectric focusing. The usual disadvantages of all of these methods are that they require the starting material be passed through several processes to remove contaminants to the point where the desired material is substantially pure.

In immunoaffinity chromatography, an antibody to the desired protein or other molecule is immobilized on a chromatographic substrate, the protein mixture is applied to the substrate under conditions allowing the antibody to bind the protein, the unbound material is removed by washing, and the bound protein is eluted using, for example, high or low pH, protein denaturants or chaotropes. The end result is a substantially pure protein which often lacks full biological activity.

A variation of this method is described in U.S. Ser. No. 07/730,040 filed Jul. 12, 1991, which is a continuation of U.S. Ser. No. 07/292,447 entitled "Monoclonal Antibody against Protein C" filed Dec. 30, 1988 by Charles T. Esmon and Naomi L. Esmon, disclosing the properties of the monoclonal antibody, HPC-4. The hybridoma cell line which secretes the monoclonal antibody designated as HPC-4, was deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md, on Nov. 2, 1988, and assigned ATCC No. HB 9892. This deposit is available to the public upon the grant of a patent.

HPC-4 binds protein C, not activated protein C (APC), and only in the presence of calcium. Thus, when the antibody is immobilized on an affinity support, protein C can be isolated from either plasma-derived sources or from tissue culture expression systems under extremely mild conditions. This is important in maintaining the biological activity of the product and the stability of the solid support resin. Because activated protein C is not bound under any conditions, the resulting product is completely free of APC.

The antibody binds to a defined region of the protein C molecule that is contained within residues 6 and 17 of the heavy chain, specifically E D Q V D P R L I D G K. This peptide can be immobilized directly on a solid support resin and can be used to isolate the antibody in high concentrations from mouse ascites fluid or tissue culture supernatants. This approach allows the isolation of the antibody in extremely pure form in high yield, even from very dilute solutions. The antibody can be removed from the solid support peptide either by the removal of calcium ions, if desired, or by 1.5 M guanidine, which does not affect the function of the purified monoclonal antibody.

It would be advantageous if the purification methods using the Protein C epitope in combination with HPC-4 could be applied to the purification of other proteins, especially blood clotting proteins One such system using a completely different antibody has been described by Prickett, et al., *BioTechniques* 7(6), 580–589 (1989), using a calcium-dependent antibody that recognizes several of the enterokinase sites that are used for Ca^{2+}coordination, including residues 1, 2, 3, 4, and 7 from FIG. 4, p. 583 and Table I, p. 586. This sequence is attached to the N-terminus of the protein to be isolated, and subsequently removed by treatment with urokinase.

It is therefore an object of the present invention to provide a method and means allowing isolation of purified proteins, especially blood clotting proteins, in a single chromatographic step.

It is a further object to provide recombinant proteins having an amino acid sequence specifically bound by a monoclonal antibody.

SUMMARY OF THE INVENTION

A method is disclosed to make any protein in a form that can be isolated rapidly from a solution using a specific monoclonal antibody designated "HPC-4". HPC-4 binds a twelve amino acid epitope of Protein C zymogen recognized in combination with calcium. It has now been determined that it is possible to form a fusion protein of the epitope with a protein to be isolated, and isolate the protein using HPC-4-based affinity chromatography. In the preferred embodiment, a specific protease cleavage site is inserted between the epitope and the protein so that the epitope can be easily removed from the isolated protein. In the most preferred embodiment, the fusion protein is formed by expression of a recombinant gene inserted into an appropriate vector.

In an example, a functionally active soluble tissue factor including the twelve amino acid epitope of Protein C zymogen recognized in combination with calcium by a specific antibody to Protein C zymogen ("HPC-4") and a factor Xa cleavage site was expressed from a vector inserted into a suitable procaryotic or eukaryotic expression system. The recombinant soluble tissue factor can be rapidly isolated in a single chromatographic step using the HPC-4 monoclonal antibody immobilized on a suitable substrate. Once isolated, the Protein C epitope is removed by cleavage with factor Xa, leaving the functionally active, soluble tissue factor.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is expression vector pIN-III-pelB encoding the HPC-4 epitope and factor Xa cleavage site, followed by several restriction enzyme sites for insertion of the gene encoding the protein to be isolated.

FIG. 2 is a schematic representation of steps involved in the construction of pIN-III-pelB-tTF expression vector. Subcloning of tTF was performed in two stages. At

the first stage an intermediary plasmid was prepared which harbored the tTF gene (pUC-tTF). At the second stage the tTF gene was removed from pUC-tTF and subcloned into the pIN-III-pelB (see the text for more detail). The resulting pIN-III-pelB-tTF was used for the transformation of *E. coli* (XL1-B) for expression.

FIG. 3 is a photograph of an SDS gel electrophoresis of the tissue factor, which runs as a monomer in the gel even without disulfide bond reduction.

FIG. 4 is a graph of Xa generation rate (m absorbance/min/min) versus tTF (nM) for TTF-HPC-4 (open circles); TTF-293 (closed circles); and TTF-Xa-DIG (open squares).

DETAILED DESCRIPTION OF THE INVENTION

A fusion protein readily isolated by affinity chromatography using HPC-4 antibody is prepared by insertion of a DNA sequence encoding the twelve amino acid HPC-4 epitope into a vector, followed by the gene encoding the protein to be isolated. In the preferred embodiment, a specific protease cleavage site is inserted into the vector between the epitope and protein coding sequence, so that the resulting fusion protein can be easily cleaved to yield the epitope peptide and the desired protein.

In the following non-limiting example, a nucleic acid sequence encoding the twelve amino acid residue epitope for HPC-4 is inserted into an expression vector in an orientation such that the expression protein contains the HPC-4 epitope followed by a factor Xa cleavage site and then the amino terminus of tissue factor which has had the signal peptide cytosolictail and transmembrane spanning domains deleted.

The HPC-4 Monoclonal Antibody

The HPC-4 antibody and uses thereof are described in U.S. Ser. No. 07/730,040 filed Jul. 12, 1991, which is a continuation of U.S. Ser. No. 07/292,447 entitled "Monoclonal Antibody against Protein C" filed Dec. 30, 1988 by Charles T. Esmon and Naomi L. Esmon, the teachings of which are incorporated herein. A detailed analysis of the properties of the HPC-4 monoclonal is presented in Stearns, et al., "The Interaction of a Ca^{2+}-Dependent Monoclonal Antibody with the Protein C Activation Peptide Region," J. Biol. Chem. 263, 826–832 (1988).

The HPC-4 monoclonal antibody is directed against a peptide sequence present in the activation region of the heavy chain of Protein C and Ca^{2+}. This peptide sequence consists of twelve amino acids, glutamic acid-aspartic acid-glutamine-valine-aspartic acid-proline-arginine-leucine-isoleucine-aspartic acid-glycine-lysine (E D Q V D P R L I D G K) (SEQUENCE 1). An advantage of this sequence is that it is short enough to be made synthetically but long enough to impart specificity, thereby avoiding potential cross reactions between the antibody and proteins other than the fusion protein.

The antibody appears to have at least one metal ion binding site in addition to the peptide binding site. The peptide binding activity is responsive to, or "dependent on", binding at the metal ion binding site. The metal ion binding site is capable of binding to a divalent metal cation such as calcium, or a metal having a similar ionic radius and coordination properties such as Tb^{3+}. The peptide does not bind Ca^{2+} and hence no Ca^{2+} binding site is added to the fusion protein. This will minimize nonspecific Ca^{2+} mediated interaction with the isolation matrix potentially inherent in the Prickett approach, described at page 3. In the case of tissue factor, it also means that calcium binding to factor VII (the ligand for tissue factor) can be studied without interference due to an additional metal binding site in the tissue factor fusion protein.

When calcium binds to the metal ion binding site in the antibody, the monoclonal antibody becomes significantly more receptive to binding to the peptide. When a metal ion is not bound to the metal ion binding site of the monoclonal antibody, the antigen binding site is relatively unreceptive to binding the antigen. Accordingly, antibody-antigen binding may be controlled by varying the metal ion concentration in the media surrounding the antibody.

Proteins to be Expressed and Purified.

The method described herein is not limited as to the protein that can be expressed as a fusion protein, isolated using HPC-4, then separated from the epitope as the pure protein.

Vectors and Expression Systems

A vector is selected for expression of a sequence encoding the pelB leader peptide HPC-4 epitope and protein to be isolated, preferably separated by a specific protease cleavage site. Examples of suitable bacterial expression vectors which are commercially available include pcDNA II (Invitrogen #V400-20), pNH8a (Stratagene, #215201) and pBTacI (Boehringer Mannheim, #1081365). These vectors are used for expression of full length or partial cDNA sequences in bacteria such as *E. coli*, where the expressed protein usually accumulates as insoluble aggregates in the cytoplasm of bacteria called inclusion bodies. To extract the target protein, a high concentration (usually 8 M) of a chaotropic agent such as urea is required to dissolve inclusion bodies. This process denatures the protein and results in the inactivation of the target protein. To obtain a functional protein, a refolding step is necessary which is usually very inefficient and difficult to control. Examples of commercially available mammalian expression vectors include pRc/RSV (Invitrogen, #V780-20), pRc/CMV (Invitrogen #V750-20), and pMC1NeO (Stratagene #213201). These expression vectors usually contain a suitable promotor that can direct high-level expression of recombinant proteins in mammalian cells and they also contain a drug resistant gene that can be used for selection of those mammalian cells that have integrated these vectors into their genomes. These vectors are suitable for the expression of full length cDNA and any other DNA fragment which contains a leader peptide at the 5' end of the sequence. These constructs are transferred into a suitable expression system, either procaryotic cells such as *E. coli*, or eukaryotic cell, such as a yeast or mammalian cell culture system.

It is also possible to insert the cDNA encoding the fusion protein into an embryo for production of a transgenic animal for production of the protein using known methodology. The protein expression can be targeted to a specific tissue using a tissue specific promotor in combination with the protein encoding sequence. For example, the fusion gene is isolated on 1% agarose gel followed by electroelution in a dialysis bag, as described by Maniatis, et al. (1982). The eluted DNA is precipitated, redissolved in water and purified by passing through an elutip-D column as per the instructions of the manufacturer (Schleicher and Schuell, Inc., Keene, NH). The purified DNA is dissolved in 5 mM Tris (pH

7.4) and 0.1 mM EDTA at 3 μg/ml concentration for microinjection.

Mice or other suitable animals such as rabbits or sheep embryos are obtained from commercial suppliers. Reagents such as bovine serum albumin, gelatin, and pronase are obtained from Sigma Chemical Co , St. Louis, MO. Hormones for superovulation, PMS and hCG, are obtained from Organon, Inc., NJ. Hyaluronidase is purchased from Sigma. Restriction enzymes are obtained from New England Biolabs, Beverly, MA. The micromanipulator made by Nara Shige, USA, Inc., Rainin Instruments Co., Woburn, MA can be used to microinject DNA into the pronuclei. DMEM, fetal bovine serum, and DPBS can be obtained from GIBCO Laboratories, Gaithersville, MD.

For construction of transgenic mice, procedures for embryo manipulation and microinjection are described in "Manipulating the Mouse Embryo" by B. Hogan, F. Costantini and E. Lacy (Cold Spring Harbor Laboratory, 1986). Similar methods are used for production of other transgenic animals. Mouse zygotes are collected from six week old females that have been superovulated with pregnant mares serum (PMS) followed 48 hours later with human chorionic gonadotropin. Primed females are placed with males and checked for vaginal plugs on the following morning. Pseudopregnant females are selected for estrus, placed with proven sterile vasectomized males and used as recipients. Zygotes are collected and cumulus cells removed by treatment with hyaluronidase (1 mg/ml). Pronuclear embryos are recovered from female mice mated to males. Females are treated with pregnant mare serum, PMS, (5 IU) to induce follicular growth and human chorionic gonadotropin, hCG (51 U) to induce ovulation. Embryos are recovered in a Dulbecco's modified phosphate buffered saline (DPBS) and maintained in Dulbecco's modified essential medium (DMEM) supplemented with 10% fetal bovine serum.

Microinjections can be performed using Narishige micromanipulators attached to a Nikon diaphot microscope. Embryos are held in 100 microliter drops of DPBS under oil while being microinjected. DNA solution is microinjected into the largest visible male pronucleus. Successful injection is monitored by swelling of the pronucleus. Immediately after injection embryos are transferred to recipient females, mature mice mated to vasectomized male mice. Recipient females are anesthetized using 2,2,2-tribromoethanol. Paralumbar incisions are made to expose the oviducts and the embryos are transformed into the ampullary region of the oviducts. The body wall is sutured and the skin closed with wound clips. Recipients are appropriately ear notched for identification and maintained until parturition.

At three weeks of age about 2–3 cm long tail samples are excised for DNA analysis. The tail samples are digested by incubating overnight at 55° C. in the presence of 0.7 ml 50 mM Tris, pH 8.0, 100 mM EDTA, 0.5% SDS and 350 μg of proteinase K. The digested material is extracted once with equal volume of phenol and once with equal volume of phenol:chloroform (1:1 mixture). The supernatants are mixed with 70 μl 3 M sodium acetate (pH 6.0) and the DNAs are precipitated by adding equal volume of 100% ethanol. The DNAs are spun down in a microfuge, washed once with 70% ethanol, dried and dissolved in 100 μL TE buffer (10 mM Tris, pH 8.0 and 1 mM EDTA). 10 to 20 μl of DNAs were cut with BamHI and Bgl11 or EcoRI, electrophoresed on 1% agarose gels, blotted onto nitrocellulose paper and hybridized with ^{32}P-labeled DNA sequences. Transgenic animals are identified by autoradiography.

The transgenic females are mated. At five days following parturition milk samples were taken and assayed for the fusion protein. At six to seven weeks of age transgenic males are mated. The Fl litters are analyzed for transgene. The positive females are kept and mated at five weeks of age. At five days following parturition milk samples are assayed for the fusion protein. Milk samples (50–200 μl) are collected from anesthetized mice injected with 0.05 units of oxytocin, an inducer of lactation. The milk is collected in a glass capillary with the aid of mammary palpation. The fusion protein is then isolated by binding to the HPC-4 antibody.

Purification using the HPC-4 antibody-epitope.

The antibody can be bound to a variety of substrates, for use in purification and isolation of the fusion protein, including agarose, acrylamide and other types of conventional chromatographic resins, filters, etc. These materials are known to those skilled in the art, as are the methods for attaching the protein to them. The selection of the material will depend in large part on the scale of the purification or the sample to be analyzed, as well as biocompatibility and government agency approval where the end-product is for pharmaceutical use.

Protease Cleavage Site.

In the most preferred embodiment, the fusion protein includes a protease cleavage site between the epitope and the protein to be isolated. Suitable sites include sequences cleaved by Factor Xa: Ile Glu Gly Arg (IEGR), enterokinase: Asp Asp Asp Asp Lys (DDDDK), and thrombin: Phe/Gly Pro Arg (F/GPR).

Following purification with the HPC-4, the fusion protein is treated with the appropriate enzyme to cleave the binding peptide from the desired protein.

The present invention will be further understood by reference to the following non-limiting examples.

EXAMPLE 1

Construction of a vector for expression of a fusion truncated tissue factor.

Blood coagulation results from the production of thrombin, a proteolytic enzyme inducing platelet aggregation and cleaving fibrinogen to fibrin, which stabilizes the platelet plug. A number of proenzymes and procofactors circulating in the blood interact in this process through several stages during which they are sequentially or simultaneously converted to the activated form, ultimately resulting in the activation of prothrombin to thrombin by activated factor X (fXa) in the presence of factor Va, ionic calcium, and platelets.

Factor X can be activated by either of two pathways, termed the extrinsic and intrinsic pathways. The intrinsic pathway, or surface-mediated activation pathway, consists of a series of reactions where a protein precursor is cleaved to form an active protease, beginning with activation of factor XII to factor XIIa, which converts factor XI to factor XIa, which, in the presence of calcium, converts factor IX to factor IXa. Factor IX can also be activated via the extrinsic pathway by tissue factor (TF) in combination with activated factor VII (factor VIIa; fVIIa). The activated factor IX, in the presence of calcium, phospholipid (platelets), and factor VIIIa, activates factor X to factor Xa.

Physiologically, the major pathway involved in coagulation is believed to be the extrinsic pathway, an essen-

tial step of which is activation of factor VII to factor VIIa. Clotting assays and other activity assays designed to measure factor VII and VIIa generally must employ TF, the cofactor required for factor VIIa coagulant activity. Most commonly, TF is provided as a relatively crude preparation known as thromboplastin. Tissue factor is an integral membrane glycoprotein having a protein and a phospholipid component. It has been isolated from a variety of tissues and species and reported to have a molecular mass of between 42,000 and 53,000. DNA encoding tissue factor and methods for expression of the protein have now been reported, for example, in European Patent Application 0 278 776 by Genentech, Inc. and by J. H. Morrissey, et al. Cell 50, 129–135 (1987).

The nucleotide (SEQUENCE 2) and amino acid (SEQUENCE 3) sequence of truncated tissue factor (tTF) is shown below, which, as described below, was modified from the sequence described in U.S. Ser. No. 07/683,682 filed Apr 10, 1991, the teachings of which are incorporated herein. The truncated tissue factor protein lacks the predicted transmembrane and cytoplasmic domains of tissue factor. The essential difference between truncated tissue factor and wild-type tissue factor is that truncated tissue factor is no longer tethered to the phospholipid membrane surface. Soluble tissue factor is a cofactor for activated factor VII (FVII) but not precursor factor VII (FVII). Intact tissue factor is a cofactor for FVII and FVIIa. The SmaI site at the 5! end, and the XbaI site at the 3' end are underlined. The first three nucleotides of the SmaI site are removed upon digestion with the restriction enzyme, SmaI. This permits the tTF cDNA sequence to be blunt-end ligated to StuI site of the pIN-III-pelB-HPC-4 expression vector in a manner that preserves the reading frame of the tTF cDNA clone.

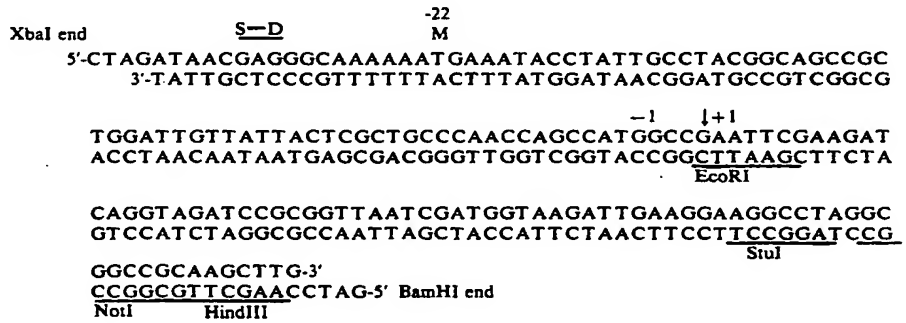
Construction of pIN-III-pelB Expression Vector:

This vector is derived from pIN-III-ompA (John Ghrayeb et al, EMBO J. 3(10):2437–2442 (1984)). Cleavage of pIN-III-ompA with XbaI and BamHI removes a DNA fragment which contains the Shine-Dalgarno sequence GAGG, and the entire nucleotide sequences encoding for the ompA signal peptide. As shown in FIG. 1, by ligation of 8 overlapping oligonucleotides (four sense and the other four complementary antisense), a DNA fragment was synthesized which contains the missing Shine-Dalgarno sequence followed by oligonucleotides sequence encoding a 22 residue long peptide representing the pelB signal peptide (Sahu-Ping Lei, et al, J. Bacteriol. 169(9):4379–4383, 1987) (AA −22 to −1) (SEQUENCE 4).

An EcoRI restriction site was included in this DNA fragment, immediately after the pelB leader sequence, which encodes two extra residues Glu, Phe (+1 and 2). The DNA fragment also encodes a 12 amino acid residue long peptide (residue 3 to 14), which is the epitope for the Ca++dependent monoclonal antibody HPC-4. The EcoRI restriction site separates the HPC-4 epitope from the pelB signal peptide. Following the nucleotides encoding the epitope, are the nucleotides encoding the four amino acid residues Ile, Glu, Gly, and Arg, forming a factor Xa cleavage site (SEQUENCE 6).

The factor Xa cleavage site is followed by a sequence containing several restriction enzyme sites for cloning of the target gene into this vector for expression. As shown in FIG. 1, this DNA fragment contains an XbaI sticky end at the 5' end and a BamHI sticky end at the 3' end for ligation to the XbaI and BamHI sites of pIN-III-ompA. Multiple cloning sites of pIN-III-pelB includes StuI, Not I, HindIII, and BamHI. StuI is a 6 bp blunt end cutter whose recognition site is AGG CCT; when it cleaves this site it leaves three nucleotides at the 3' end of the DNA fragment, AGG, which encode the last residue in the FXa cleavage site, Arg. Therefore, double digestion of this vector with StuI and either one of the other 3' end cloning sites (NotI or HindIII or BamHI) provides a suitable directional cloning of the target gene for expression.

The oligonucleotide sequence of a synthetic DNA fragment encoding the pelB leader peptide (AA-22 to −1) (SEQUENCE 4); HPC-4 epitope (AA +3 to 14) (SEQUENCE 1); and the FXa cleavage site (AA 15 to 18) (SEQUENCE 6), are shown below. The Shine-Dalgarno sequence GAGG (S-D) is overlined. The restriction enzyme sites useful for cloning have been outlined, which are EcoRI (between the pelB leader sequence and the HPC-4 epitope which creates two extra amino acids +1 and +2), StuI, NotI, and HindIII. The pelB leader peptide cleavage site which is cleaved by bacterial signal peptidase is shown by an arrow. This DNA fragment was made by ligation of eight overlapping oligonucleotides so that it created a sticky XbaI site at the 5' end and a sticky BamHI site at the 3' end for ligation into the XbaI and BamHI site of the pIN-III-ompA vector. The boundary of oligonucleotides for the sense strand and at the bottom for the complementary antisense strand is shown by bold letters.



Construction of pIN-III-pelB-tTF

The construction of pIN-III-pelB-tTF is shown schematically in FIG. 2. pJH27 is a plasmid that contains the entire truncated tissue factor (tTF) cDNA sequence cloned between BamHI and XbaI restriction sites of the vector pGEM-7Zf(+) (Promega). Preparation of the tTF cDNA fragment with 5' and 3' ends compatible with cloning sites in the pIN-III-pelB expression vector

was performed in a two-step process. At the first stage the plasmid pJH27 was digested with BbvI restriction enzyme which cleaved away DNA sequences from the 5' end of the cDNA coding for the tTF signal peptide and the first 11 residues of the mature tTF protein. In order to repair the tTF gene two complementary oligonucleotides were synthesized, which coded for the 11 missing residues and created a blunt, half-SmaI site at the 5' end. This oligonucleotide contained a sticky end that could hybridize to, and therefore be ligated to, the sticky end of the tTF cDNA sequence created by BbvI digestion.

In the second stage, the modified cDNA insert was removed from pJH27 by digestion with XbaI, and the insert was resolved by agarose gel electrophoresis and purified by elution from the agarose gel. The resulting cDNA fragment which encodes the entire tTF was subcloned into the SmaI and XbaI sites of pUC19 plasmid. As shown in FIG. 2, the tTF cDNA fragment then was removed from pUC19 with SmaI and HindIII restriction enzymes and subcloned into the StuI and HindIII sites of the pIN-III-pelB expression vector.

EXAMPLE 2

Expression and Isolation of the Fusion Protein

Growth of Bacteria for Periplasmic Extract The XL1-B strain of *E. coli* was grown at 37° C. in LB media. 500 μl of overnight culture was transferred to a 125 ml flask containing 25 ml of LB and 100 μg/ml ampicillin. The flask was shaken at 37° C. for 2–3 hrs until the OD_{600} is equal to 0.6 to 0.7. One liter LB media with ampicillin (Amp) was inoculated with the entire 25 ml bacterial culture and the shaking was continued at 37° C. until the OD_{600}=0.6 to 0.7. The culture then was induced with 1 mM isopropyl-B-D-thiogalactopyranoside (IPTG) and the shaking was continued for another 8 hrs at room temperature. The bacterial culture was centrifuged at approximately 3000 rpm for 20 min to separate the culture medium from the cells. The cell pellet was resuspended with 50 ml of cold water and incubated for ½ hour on ice with shaking. The periplasmic extract was then collected by spinning at 10,000 g for ½ hr.

Approximately three-fourths of the protein was expressed in the media of the *E. coli*, and one-fourth was recovered from the periplasmic space after hypotonic shock. The periplasmic extract was mixed with the culture medium (which has been clarified by centrifugation, as described above), and the mixture was concentrated to approximately one-tenth the original volume using an ultrafiltration spiral cartridge concentrator with a 3000 MW cut-off membrane (Amicon). Following concentration, this material was brought up to 0.1 M NaCl, 0.02 M Tris and 1 mM calcium chloride, and loaded on a HPC-4 column equilibrated with the same buffer (5 mg/ml HPC-4 IgG immobilized on Affi-Gel™ 10; 4 ml total used for 1 liter of starting bacterial culture), washed with approximately 200 ml 1 M NaCl, 0.02 M Tris-HCl, pH 7.5 containing 1 mM $CaCl_2$, followed by approximately 10 to 20 ml of the same buffer but with 0.1 M NaCl, and the protein eluted with 0.1 M NaCl, 0.02 M Tris-HCl, 5 mM EDTA, pH 7.5. Elution of the tTF from the column was monitored by absorbance of light at 280 nm; a single peak of protein was observed.

SDS-PAGE

SDS-PAGE analysis (10% acrylamide using Laemmli system, Laemmli, Nature 227:680–685 (1970)) of the peak fraction from the affinity chromatography indicated a single monomeric band at around 30 kd in both reducing and non-reducing conditions, as shown in FIG. 3. This is consistent with the expected molecular weight of tTF when it is not glycosylated.

Approximately 1 mg of soluble tissue factor is recovered from 1 liter of starting bacterial culture, and is homogeneous as shown by SDS gel electrophoresis. The tissue factor runs as monomer in the gel even without disulfide bond reduction. The soluble, truncated tissue factor protein isolated in this manner has full cofactor activity toward factor VIIa even without the HPC-4 epitope removed; this activity is equivalent to that of truncated tissue factor without the HPC-4 epitope expressed in mammalian cells, as shown by FIG. 4.

Functionally the protein is equivalent to the protein expressed in mammalian cell culture, but at less than 10% the cost. This functional equivalence is demonstrated by the factor VIIa concentration dependence on factor X activation and by the ability of the tissue factor from both sources to increase factor VIIa amidolytic activity equivalently.

EXAMPLE 3

Isolation of other proteins using the fusion protein technology in combination with the HPC-4 epitope.

Thrombomodulin fragments, EGF domains of factor X and EGF domains of protein C with the epitope linked to the amino terminal region of these proteins have been expressed in and recovered from other *E. coli* periplasmic space expression systems. A fragment of thrombomedulin of the 4th-14 6th EGF domains has also been isolated from cultured mammalian cells in a single step purification on the antibody column as described above. In this case, the DNA encoding the thrombomedulin 4th–6th EGF domains was amplified from cDNA by PCR and ligated to synthetic oligonucleotides coding for the transferrin signal peptide followed by the HPC-4 epitope and Xa cleavage site. The resulting DNA fragment was subcloned into a pRc/RSV mammalian cell expression vector for expression in human 293 cells.

Modifications and variations of the method and compositions described herein will be obvious to those skilled in the art from the foregoing detailed description. Such modifications and variations are intended to come within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 9

(2) INFORMATION FOR SEQ ID NO:1:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Mouse

(i x) FEATURE:
(A) NAME/KEY: Binding-site
(B) LOCATION: 1..12
(D) OTHER INFORMATION: /note="Epitope recognized by HPC4 antiprotein C antibody"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Glu Asp Gln Val Asp Pro Arg Leu Ile Asp Gly Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 672 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(i x) FEATURE:
(A) NAME/KEY: miscrecomb
(B) LOCATION: 1..6

(i x) FEATURE:
(A) NAME/KEY: miscrecomb
(B) LOCATION: 67..72

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCCGGGACCA CCAATACTGT GGCAGCATAT AATTTAACTT GGAAATCAAC TAATTTCAAG 60
ACAATTTTGG AGTGGGAACC CAAACCCGTC AATCAAGTCT ACACTGTTCA AATAAGCACT 120
AAGTCAGGAG ATTGGAAAAG CAAATGCTTT TACACAACAG ACACAGAGTG TGACCTCACC 180
GACGAGATTG TGAAGGATGT GAAGCAGACG TACTTGGCAC GGGTCTTCTC CTACCCGGCA 240
GGGAATGTGG AGAGCACCGG TTCTGCTGGG GAGCCTCTGT ATGAGAACTC CCCAGAGTTC 300
ACACCTTACC TGGAGACAAA CCTCGGACAG CCAACAATTC AGAGTTTTGA ACAGGTGGGA 360
ACAAAAGTGA ATGTGACCGT AGAAGATGAA CGGACTTTAG TCAGAAGGAA CAACACTTTC 420
CTAAGCCTCC GGGATGTTTT TGGCAAGGAC TTAATTTATA CACTTTATTA TTGGAAATCT 480
TCAAGTTCAG GAAAGAAAAC AGCCAAAACA AACACTAATG AGTTTTTGAT TGATGTGGAT 540
AAAGGAGAAA ACTACTGTTT CAGTGTTCAA GCAGTGATTC CCTCCCGAAC AGTTAACCGG 600
AAGAGTACAG ACAGCCCGGT AGAGTGTATG GGCCAGGAGA AAGGGGAATT TAGAGAATAA 660
CTGCAGTCTA GA 672

(2) INFORMATION FOR SEQ ID NO:3:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gly	Thr	Thr	Asn	Thr	Val	Ala	Ala	Tyr	Asn	Leu	Thr	Trp	Lys	Ser	Thr
1				5					10					15	
Asn	Phe	Lys	Thr	Ile	Leu	Glu	Trp	Glu	Pro	Lys	Pro	Val	Asn	Gln	Val
			20					25					30		
Tyr	Thr	Val	Gln	Ile	Ser	Thr	Lys	Ser	Gly	Asp	Trp	Lys	Ser	Lys	Cys
		35					40					45			
Phe	Tyr	Thr	Thr	Asp	Thr	Glu	Cys	Asp	Leu	Thr	Asp	Glu	Ile	Val	Lys
	50					55					60				
Asp	Val	Lys	Gln	Thr	Tyr	Leu	Ala	Arg	Val	Phe	Ser	Tyr	Pro	Ala	Gly
65					70					75					80
Asn	Val	Glu	Ser	Thr	Gly	Ser	Ala	Gly	Glu	Pro	Leu	Tyr	Glu	Asn	Ser
				85					90					95	
Pro	Glu	Phe	Thr	Pro	Tyr	Leu	Glu	Thr	Asn	Leu	Gly	Gln	Pro	Thr	Ile
			100					105					110		
Gln	Ser	Phe	Glu	Gln	Val	Gly	Thr	Lys	Val	Asn	Val	Thr	Val	Glu	Asp
		115					120					125			
Glu	Arg	Thr	Leu	Val	Arg	Arg	Asn	Asn	Thr	Phe	Leu	Ser	Leu	Arg	Asp
	130					135					140				
Val	Phe	Gly	Lys	Asp	Leu	Ile	Tyr	Thr	Leu	Tyr	Tyr	Trp	Lys	Ser	Ser
145					150					155					160
Ser	Ser	Gly	Lys	Lys	Thr	Ala	Lys	Thr	Asn	Thr	Asn	Glu	Phe	Leu	Ile
				165					170					175	
Asp	Val	Asp	Lys	Gly	Glu	Asn	Tyr	Cys	Phe	Ser	Val	Gln	Ala	Val	Ile
			180					185					190		
Pro	Ser	Arg	Thr	Val	Asn	Arg	Lys	Ser	Thr	Asp	Ser	Pro	Val	Glu	Cys
		195					200					205			
Met	Gly	Gln	Glu	Lys	Gly	Glu	Phe	Arg	Glu						
	210					215									

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(i x) FEATURE:
(A) NAME/KEY: miscfeature

-continued

(B) LOCATION: 10..13
(D) OTHER INFORMATION: /note="Shine-Dalgarno sequence"

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 22..85
(D) OTHER INFORMATION: /note="pelB leader peptide"

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 3..14
(D) OTHER INFORMATION: /note="HPC-4 epitope"

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 15..18
(D) OTHER INFORMATION: /note="Factor Xa cleavage site"

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 146..152
(D) OTHER INFORMATION: /note="NotI restriction enzyme"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
CTAGATAACG AGGGCAAAAA ATGAAATACC TATTGCCTAC GGCAGCCGCT GGATTGTTAT      60
TACTCGCTGC CCAACCAGCC ATGGCCGAAT TCGAAGATCA GGTAGATCCG CGGTTAATCG     120
ATGGTAAGAT TGAAGGAAGG CCTAGGCGGC CGCAAGCTTG                           160
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 83..88
(D) OTHER INFORMATION: /note="EcoRI restriction enzyme site"

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 134..139
(D) OTHER INFORMATION: /note="StuI restriction enzyme site"

(i x) FEATURE:
(A) NAME/KEY: miscfeature
(B) LOCATION: 150..155
(D) OTHER INFORMATION: /note="HindIII restriction enzyme"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
TATTGCTCCC GTTTTTTACT TTATGGATAA CGGATGCCGT CGGCGACCTA ACAATAATGA      60
GCGACGGGTT GGTCGGTACC GGCTTAAGCT TCTAGTCCAT CTAGGCGCCA ATTAGCTACC     120
ATTCTAACTT CCTTCCGGAT CCGCCGGCGT TCGAACCTAG                           160
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(i x) FEATURE:
(A) NAME/KEY: Cleavage-site
(B) LOCATION: 1..4
(D) OTHER INFORMATION: /note="Factor Xa Cleavage Site"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ile Glu Gly Arg
1

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(i x) FEATURE:
(A) NAME/KEY: Cleavage-site
(B) LOCATION: 1..5
(D) OTHER INFORMATION: /note="Enterokinase Cleavage Site"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asp Asp Asp Asp Lys
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(i x) FEATURE:
(A) NAME/KEY: Cleavage-site
(B) LOCATION: 1..3
(D) OTHER INFORMATION: /note="Thrombin Cleavage Site"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Phe Pro Arg
1

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

-continued

(i x) FEATURE:
(A) NAME/KEY: Cleavage-site
(B) LOCATION: 1..3
(D) OTHER INFORMATION: /note= "Thrombin Cleavage Site"

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Gly Pro Arg
1

We claim:

1. A fusion protein comprising an epitope which is capable of binding in the presence of calcium to the monoclonal antibody designated as HPC-4 (ATCC No. HB 9892) and a protein to be isolated by binding of the fusion protein to HPC-4 antibody, wherein the epitope and the protein to be isolated are separated by a specific protease cleavage site.

2. The fusion protein of claim 1 wherein the epitope consists of the amino acid sequence E D Q V D P R L I D G K (SEQ ID NO: 1).

3. The fusion protein of claim 1 wherein the protease cleavage site is selected from the group consisting of amino acid sequences specifically cleaved by Factor Xa; Ile Glu Gly Arg (IEGR, SEQ ID NO; 6), enterokinase: Asp Asp Asp Asp Lys (DDDDK, SEQ ID NO: 7), and thrombin: Phe/Gly Pro Arg (F/GPR, SEQ ID NO: 8 and SEQ ID NO: 9).

4. The fusion protein of claim 1 wherein the protein to be isolated is selected from the group consisting of tissue factor and a protein corresponding to the fourth through sixth epidermal growth factor domains of thrombomodulin.

5. The fusion protein of claim 1 consisting essentially of the amino acid sequence E D Q V D P R L I D G K (SEQ ID NO: 1) at the N-terminus, an amino acid sequence specifically cleaved by factor Xa, and tissue factor.

6. The fusion protein of claim 5 wherein the tissue factor is the extracellular domain of tissue factor.

* * * * *

Protein Expression and Purification 15, 228–235 (1999)
Article ID prep.1998.1015, available online at http://www.idealibrary.com on IDEAL®

Overexpression, Oxidative Refolding, and Zinc Binding of Recombinant Forms of the Murine S100 Protein MRP14 (S100A9)

Mark J. Raftery,[1] Lindsay Collinson, and Carolyn L. Geczy
Cytokine Research Unit, School of Pathology, University of New South Wales, Kensington, New South Wales, 2052, Australia

Received September 10, 1998, and in revised form November 16, 1998

Recombinant murine MRP14 (mMRP14) was produced in *Escherichia coli* using the pGEX expression system. The mass of fusion protein, by electrospray ionization–mass spectrometry (ESI/MS), was 39,213 Da which compares well with the theoretical mass (39,210.4 Da). Thrombin digestion of fusion protein was expected at a cloned thrombin consensus sequence (. . . LVPRGS . . .) located between glutathione *S*-transferase and mMRP14. Analysis of products of digestion by C4 reverse-phase HPLC and SDS–PAGE/Western blotting revealed two immunoreactive cleavage products with molecular weights around 13,000. Masses of the two proteins determined by ESI/MS were 13,062 and 11,919 Da. The larger product corresponded to the expected mass of recombinant mMRP14 (13,061.9 Da). Analysis of the protein sequence of recombinant mMRP14 revealed a thrombin-like consensus sequence (. . . NNPRGH . . .) located close to the C-terminus. The smaller protein corresponded to a truncated form of rec mMRP14 (rec MRP14$_{1-102}$) with a calculated mass of 11,918.6 Da. Optimization of the cleavage conditions resulted in >95% full-length rec mMRP14. Native mMRP14 contains one intramolecular disulfide bond between Cys$_{79}$ and Cys$_{90}$. The full-length recombinant protein was renatured and oxidized in ammonium acetate (pH ~ 7) for 96 h and formed >95% of the native intramolecular disulfide-bonded form. MRP14$_{1-102}$ bound substantially less $^{65}Zn^{2+}$ compared to native mMRP14 or rec mMRP14 after transfer to polyvinylidene difluoride and incubation with $^{65}ZnCl_2$, implicating the His residues located within the C-terminal domain in Zn^{2+} binding. © 1999 Academic Press

[1] To whom correspondence should be addressed. Fax: 61-2-9385-1389. E-mail: m.raftery@unsw.edu.au.

MRP14 [migration inhibitory factor (MIF)-related protein, 14 kDa], also known as p14, L1 heavy chain, calgranulin B, and S100A9 (1), is a small acidic protein containing two Ca^{2+}-binding EF hands belonging to the highly conserved S100 protein family (2,3). Some members also bind zinc with high affinity (1). They may influence diverse cellular processes including cell cycle progression, cell differentiation, regulation of protein phosphorylation, and cytoskeletal–membrane interactions (4,5). The C-terminal region of some members of the S100 family is extended. A peptide with sequence identity to the C-terminal domain of human MRP14 (hMRP14) has potential function (6). This region is identical to part of the sequence of neutrophil-immobilizing factor (NIF-1) which inhibits neutrophil chemotaxis to C5a and fMetLeuPro (FMLP) (6,7), but the precise role of hMRP14 in leukocyte migration is unclear. Recent experiments indicate a role in leukocyte adhesion (8). Human MRP14 has antimicrobial activity, possibly caused by its ability to sequester Zn^{2+} from the surrounding environment (9,10). A potential zinc-binding domain, located near the C-terminus of hMRP14, was identified (11).

We previously isolated CP10 (chemotactic protein, 10 kDa) and mMRP14 from supernatants of activated murine spleen cells (12,13). ESI/MS and Edman sequencing indicated that CP10 was not posttranslationally modified, whereas mMRP14 had several modifications including a disulfide bond between Cys$_{79}$ and Cys$_{90}$. The protein contains an additional free Cys residue at position 110 (13). Functional recombinant CP10 was produced using the pGEX expression system (14,15).

Formation of disulfide bonds in numerous recombinant proteins at identical Cys residues to those identified within the native protein is essential for biological activity (16–18). The importance of disulfide bonds

1046-5928/99 $30.00

for functional activity in S100 proteins is indicated for S100β, where mitogenic activity is dependent on a specific intramolecular disulfide bond between Cys_{68} and Cys_{84} (19).

To facilitate biochemical, structural, and functional characterization, a large-scale source of mMRP14 was obtained using the pGEX recombinant protein expression system. Here we describe the isolation and characterization of two forms of the protein and the oxidative refolding to produce recombinant mMRP14 with the intramolecular disulfide bridge in the same position as in the native protein. A comparison of the affinity of native mMRP14, rec mMRP14, and rec $MRP14_{1-102}$ for Zn^{2+} was also performed.

MATERIALS AND METHODS

General. Reagents and chemicals were analytical grade (Sigma, St. Louis, MO; Bio-Rad, Hercules, CA) and solvents were HPLC grade (Mallinckrodt, Clayton South, Victoria, Australia). SDS–PAGE/Western blotting were performed using a Mini Protean II apparatus (Bio-Rad) with 15% gels and a Tris/Tricine buffer system (20). Native mMRP14 was obtained from activated spleen cell supernatants as previously described (13). Liquid chromatographic separations were performed using a nonmetallic LC626 HPLC system (Waters, Bedford, MA) and UV absorbance was monitored at 214 and 280 nm with a Waters 996 photodiode array detector or 490 UV/visible detector.

Overexpression and characterization of recombinant mMRP14. The coding sequence for mMRP14 (21) was amplified by PCR from a cDNA plasmid clone, pmMRP14, and cloned between *Bam*HI and *Eco*RI sites of the glutathione *S*-transferase expression vector pGEX2T according to the manufacturer's instructions (Pharmacia Biotech, Uppsala, Sweden). The mMRP14 expression vector, pLC14, was used to transform competent *E. coli* (strain BL-21). Transformed bacteria were cultured with IPTG (100 μM) for 3–4 h to induce fusion protein production (14,15). Recombinant mMRP14 was isolated after cell lysis (3× freeze/thaw) in TBS/T (25 mM Tris, 250 mM NaCl, pH 7.5, 1% Triton X-100) using glutathione–agarose beads (Sigma) and cleaved with thrombin (Calbiochem, San Diego, CA; 20 NIH units/liter culture) in TBS (4 ml, 60 min, 37°C). Beads were washed with TBS (2 × 2 ml), the combined eluate (500 μl) was applied to a C4 RP-HPLC column (5 μm, 300 Å, Vydac, Hesperia, CA), and proteins were eluted in a linear gradient of 25 to 70% actetonitrile (0.1% TFA) at 1 ml/min over 30 min. Proteins were collected manually and concentrated using a SpeedVac (Savant, Farmingdale, NY) to a final concentration of approximately 50 ng/μl. Samples (approximately 200 ng) were analyzed by SDS–PAGE; gels were blotted onto polyvinylidene difluoride (Immobilon P, Millipore, Bedford, MA) at 75 V for 25 min before Western analysis. Semipreparative RP-HPLC was performed after reduction (5 mM DTT) and filtration (2 μm, Durapore, Millipore). The lysate (4 ml) was loaded onto a C8 RP-HPLC column (Prep-10, 20 μm, 10 × 100 mm, Brownlee/ABI, Alltech, IL) and rec mMRP14 was isolated after application of a 35 to 50% acetonitrile (0.1% TFA) gradient over 20 min at 3 ml/min. Recombinant $MRP14_{1-102}$ was isolated after prolonged (24 h) exposure to thrombin (Calbiochem; 50 NIH units/liter culture) in TBS (4 ml) at 37°C. The beads were washed with TBS (2 × 1 ml) and combined and then the truncated protein was isolated as above.

Refolding of recombinant mMRP14. Recombinant mMRP14 (~250 μg, 3 ml) isolated by preparative C8 RP-HPLC was reduced to approximately 50% of the initial volume and then dialyzed for 96 h against ammonium acetate (5 mM, pH 7, 2 liters) using 2000 MWCO Slide-a-Lyser dialysis cassettes (Pierce, Rockford, IL). The extent of refolding was assessed by C4 RP-HPLC (5 μm, 4.6 × 250 mm, Vydac) using a linear gradient of 25 to 70% acetonitrile (0.1% TFA) over 30 min at 1 ml/min at ~24-h intervals. Samples collected from C4 RP-HPLC were analyzed by SDS–PAGE and ESI/MS.

Carboxyamidomethylation of MRP14. MRP14 (5 nmol) either native, reduced, or oxidized recombinant isolated by C4 RP-HPLC were treated with iodoacetamide (19 μg, 100 nmol) in ammonium bicarbonate (100 mM, 250 μl) in the presence of guanidine–HCl (4 M) for 30 min at 30°C. After acidification (1% TFA), the mixture was loaded onto the C4 RP column and proteins were eluted with a gradient of 25 to 70% acetonitrile (0.1% TFA) at 1 ml/min over 30 min. The derivatized proteins were collected manually.

Mass spectrometry. ESI mass spectra were acquired using a single-quadrupole mass spectrometer equipped with an electrospray ionization source (Platform, VG-Fisons Instruments, Manchester, UK). Samples (~50 pmol, 10 μl) were injected into a moving solvent (10 μl/min; 50:50 water:acetonitrile, 0.05% TFA) using a Phoenix 40 HPLC pump (VG-Fisons Instruments) coupled directly to the ionization source via a fused silica capillary (50 μm × 40 cm). The source temperature was 50°C and nitrogen was used as the nebulizer and drying gas. Sample droplets were ionized at a positive potential of approximately 3 kV and transferred to the mass analyzer with a cone voltage (sample cone to skimmer lens voltage) of 50 or 70 V. The peak width at half height was 1 Da. Spectra were acquired in multichannel acquisition mode over the mass range 700 to 1800 Da in 5 s and then calibrated with horse heart myoglobin (Sigma).

Peptide mapping. Proteins (50 μg) isolated from C4 RP-HPLC were digested in ammonium bicarbonate

(250 μl, 50 mM, pH 8.0) using endoprotease Glu C (sequencing grade, Boehringer-Mannheim, Castle Hill, NSW) at an enzyme to substrate ratio of approximately 1:100 at 37°C for 6 h. The pH of the digest was lowered to approximately 2 (1% TFA) and the mixture was applied directly to a C18 RP column (Waters, Novapak, 60 Å, 4 μm, 150 × 3.9 mm). Peptides were eluted with a gradient of 5 to 75% acetonitrile (0.1% TFA) at 1 ml/min over 30 min. Peptides were collected manually and lyophilized to dryness, and water:acetonitrile (1:1, 50 μl) was then added before ESI/MS analysis.

Zinc overlay. Native mMRP14, rec mMRP14, and rec mMRP14$_{1-102}$ (~500 ng) were separated by SDS–PAGE. Proteins were transferred to PVDF at 75 V for 25 min at 4°C. After transfer, membranes were incubated in TBS or TBS + $CaCl_2$ (1 mM; 15 ml) containing $^{65}Zn^{2+}$ (25 μCi, 10 mCi/mmol, Amersham) and solutions were agitated for 30 min. Membranes were washed with TBS (3 × 15 ml) for 5 min and radioactivity was located using a phosphorimager and BI screen (GS-525, Bio-Rad). Images were analyzed using the Multianalyst program (Bio-Rad). Proteins were quantitated using a densitometer (Bio-Rad) after staining membranes with Amido black (Bio-Rad) according to the manufacturer's procedure.

RESULTS AND DISCUSSION

The pGEX expression system produces soluble recombinant proteins as a fusion with glutathione *S*-transferase (GST), enabling isolation from bacterial lysates by affinity chromatography under nondenaturing conditions. The fusion protein is digested at a cloned consensus site with either thrombin or factor Xa producing GST and recombinant protein with two additional amino acids (GlySer) at the N-terminus (22). Bacterial recombinant proteins are generally produced without disulfide bonds and must be refolded/oxidized to achieve native structures (23).

ESI/MS is a rapid and precise method for determining the mass of proteins and peptides. It can validate protein sequences (24,25) and has been used to characterize many recombinant proteins, including the S100 proteins calvasculin (26), calcyclin (26), and S100A3 (27). In these examples the proteins were expressed in *E. coli* as full-length proteins, the identity of each confirmed by ESI/MS after comparison of theoretical and experimental masses. No mutant or posttranslationally modified forms were identified. Recombinant CP10 was produced using the pGEX expression system and a mutant form containing an additional 10 C-terminal amino acids isolated and characterized using ESI/MS (15).

TABLE 1

ESI Masses of Recombinant Proteins and Calculated Masses of Proteins Derived from Proposed cDNA Sequences

Protein	C4 retention time (min)	Calculated[a] cDNA mass	ESI mass (Da)
Fusion protein	22.6	39,210.4	39,312[b]
GST	22.5	26,166.6	26,164[b]
MRP14$_{1-102}$	21.1	11,918.6	11,919
MRP14$_{reduced}$	20.5	13,061.9	13,062[c]
MRP14$_{oxidized}$	19.4	13,059.9	13,061[c]

[a] Masses were calculated from protein sequences derived from previously determined cDNA sequences.

[b] A small peak (~20%) 131 Da less was also observed, possibly due to removal of the initiator Met.

[c] Differentiation between reduced and oxidized forms is not possible because the mass difference corresponds to the approximate mass error of ±2 Da.

Isolation and ESI/MS of Recombinant mMRP14

The coding nucleotide sequence of the rec mMRP14 expression plasmid was verified by automated nucleotide sequencing (not shown). After bacterial lysis and isolation of the fusion protein using glutathione–agarose, the protein was eluted with excess glutathione and isolated after C4 RP-HPLC. The mass of fusion protein by ESI/MS was 39,213 Da, which compares favorably with the theoretical mass calculated from the fusion protein sequence derived from the constructed GST–mMRP14 cDNA (Table 1). The minor peak of 131 Da less than the full-length fusion protein was probably due to modification of the fusion protein by removal of the initiator Met (Fig. 1A). Glutathione *S*-transferase isolated after thrombin cleavage of fusion protein immobilized on GSH–agarose for 24 h had a mass of 26,164 Da (Fig. 1B) which also compared favorably with the theoretical mass (Table 1). A protein with a mass of 131 Da less than GST was also resolved (Fig. 1B) indicating a modification within the N-terminal portion of the fusion protein and supporting the proposal that the initiating Met had been removed.

After thrombin cleavage for 3 h, two proteins with retention times of 19.4 and 21.1 min were observed by analytical C4 RP-HPLC. SDS–PAGE indicated that both had molecular weights of ~13,000 Da and were immunoreactive with polyclonal antibodies to native mMRP14 by Western blotting (not shown). Only a single peak at 21.1 min was observed after prolonged cleavage ($T \sim 24$ h), whereas digestion for 1 h yielded essentially one component that eluted at 19.4 min. The mass of the early eluting peak was 13,062 Da (Fig. 2A) which corresponds to the expected theoretical mass of rec mMRP14 (Table 1) calculated from the derived protein sequence. The mass of the later-eluting protein was 11,919 Da (Fig. 2B). Analysis of the amino acid

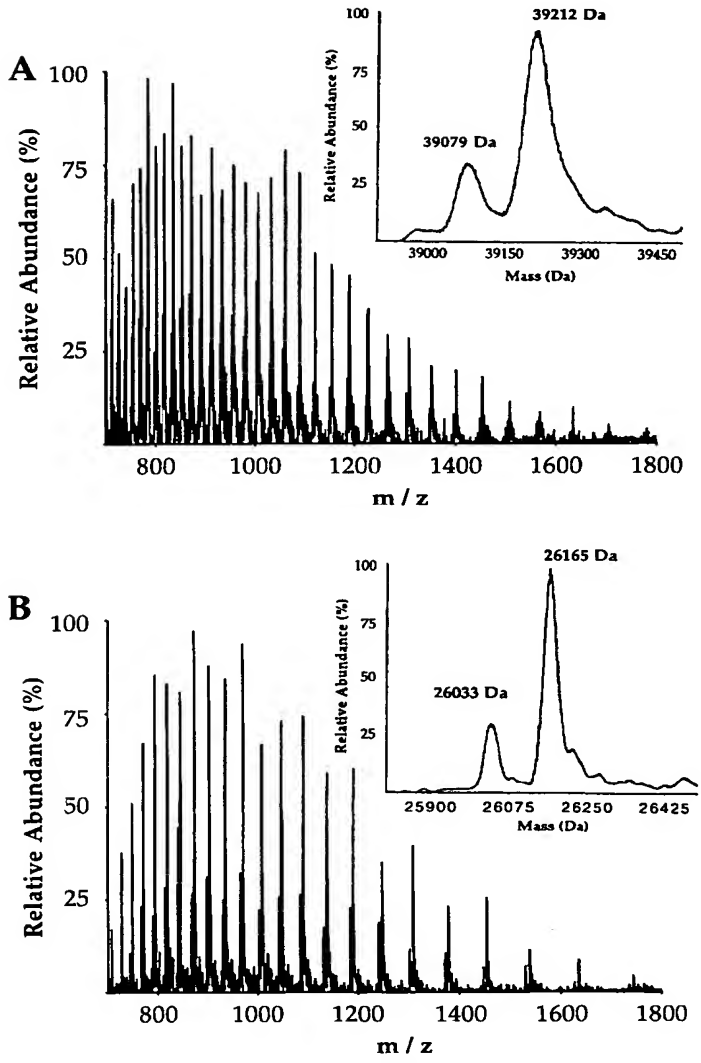


FIG. 1. ESI mass spectra showing multiply charged protein ions of rec mMRP14 fusion protein (A) and GST derived from thrombin cleavage of fusion protein (B). Deconvoluted masses are shown in insets.

sequence indicated that this corresponded to the expected mass of rec mMRP14$_{1-102}$. The amino acid sequence (. . . KLHENNPRGHGH . . .) contains a thrombin-like consensus sequence, NNPRGH, indicating that the truncated form was produced after digestion at this site. Optimum cleavage sites for thrombin have structures P4-P3-Pro-Arg-P1'-P2', where P3 and P4 are hydrophobic amino acids and P'1 and P'2 are nonacidic amino acids or P2-Arg-P1', where P2 and P1' are Gly (28). Both the engineered and native cleavage sites within the mMRP14 fusion protein contain the Pro-Arg-Gly consensus sequence. The difference in rate may be attributed to nonhydrophobic residues (Asn-Asn) located in the P4 and P3 positions of site 2, whereas site 1 contains the hydrophobic amino acids Leu-Val at P4 and P3, which is optimal for cleavage. Almost quantitative digestion of fusion protein was achieved after 1 h to yield 1.5 mg of >95% full-length recMRP14. Semipreparative C8 RP-HPLC removed residual MRP14$_{1-102}$.

Oxidative Refolding of rec mMRP14

Recombinant proteins produced in *E. coli* are often incorrectly folded, without appropriate disulfide bond formation, and normally form insoluble aggregates (inclusion bodies) during high levels of transcription (17).

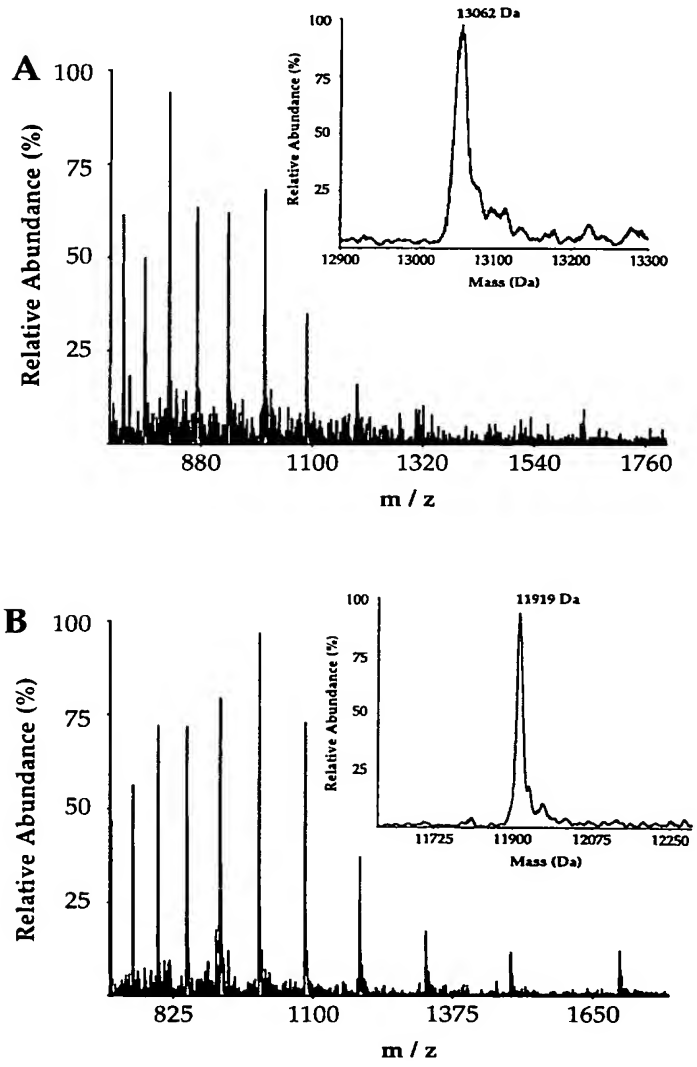


FIG. 2. ESI mass spectra showing multiply charged protein ions of the two forms of rec MRP14 isolated after C4 RP-HPLC. Full-length rec mMRP14 (A) and mMRP14$_{1-102}$ (B) derived from thrombin cleavage of fusion protein for 1 or 24 h, respectively. Deconvoluted masses are shown in insets.

Soluble recombinant proteins without disulfide bonds are also produced. Conditions for formation of native structures must be determined empirically for each protein; however, renaturation and oxidation may occur concomitantly with formation of specific disulfide bonds following solubilization/denaturation and renaturation (17,29). Generally, correctly folded native structures form *in vitro* after oxidative refolding following solubilization in urea or guanidine–HCl followed by dilution or dialysis. Proteins do undergo spontaneous oxidation in air after dilution or dialysis to yield intramolecular disulfide bonds. The rate of oxidation can be assisted by addition of mixtures of oxidized and reduced low-molecular-weight thiols (e.g. ,GSSG/GSH or cysteine/cystine) (29). Cu^{2+}, salts, detergents, and elevated temperatures can also facilitate refolding/disulfide bond formation (17). These reagents allow accelerated formation/interconversion of unstable folded intermediates and subsequent formation of thermodynamically stable disulfide bridged products, which normally represent biologically active conformations (29).

Native mMRP14 has three Cys residues near the C-terminus at positions 79, 90, and 110. We showed previously that native mMRP14 (isolated from activated spleen cell supernatants) contains a specific intramolecular disulfide bond located between Cys_{79} and Cys_{90} (13). Due to changes in relative hydrophobicity, the C4 RP-HPLC retention times of native and reduced mMRP14 differed by ~60 s facilitating characterization of the disulfide bonded form (13). Murine MRP14 is a cytosolic neutrophil protein and would be expected to be in a reduced form intracellularly. The disulfide-bridged form may be the preferred extracellular structure and may represent the functional form. Other S100 proteins are physiologically activated by oxidation. Oxidized S100β, a glial cell growth factor, contains an intramolecular disulfide bond between Cys_{68} and Cys_{84} and is the most potent S100β species to stimulate C6 glial cell proliferation (19). Oxidation of this protein also causes a conformational change which allows phosphorylation of a canonical CKII-α site located within the C-terminal Ca^{2+} binding domain, which may be functionally significant (19).

Recombinant mMRP14 was refolded and oxidized by prolonged dialysis at neutral pH. Figure 3 shows the time course of oxidative refolding of rec mMRP14 analyzed by C4 RP-HPLC. After 24 h, three different forms were identified, corresponding to oxidized mMRP14 (Fig. 3B, peak 1), the fully reduced protein (Fig. 3B, peak 2), and a transient folding intermediate (Fig. 3B, peak 3) of unknown structure. Analysis by SDS–PAGE, with silver staining and Western blotting, and ESI/MS indicated that the latter contained mixtures of monomeric and dimeric mMRP14 (not shown). The intensities of peaks 2 and 3 were reduced by prolonged dialysis until at 96 h the oxidized form (peak 1) represented the major structural form (Fig. 3D). The yield of peak 1 did not increase with more prolonged dialysis. The mass of peak 1 was identical to that expected for oxidized MRP14 (13061 Da) and the yield after semipreparative C8 RP-HPLC was ~55%. The results from each step are summarized in Table 2. The mass of the iodoacetamide derivative of oxidized rec mMRP14 was 13,117 Da. This corresponds to derivatization of a single free sulfhydryl (+57 Da) indicating that oxidized rec MRP14 contains one internal disulfide bond.

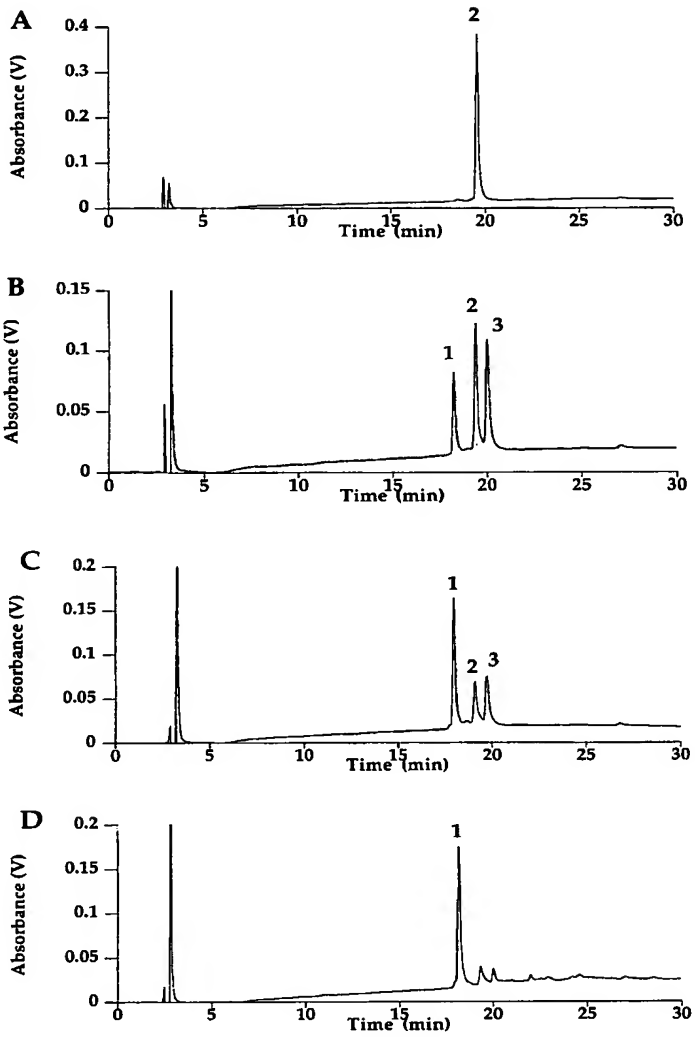


FIG. 3. RP-HPLC trace of full-length rec mMRP14 with time-dependent conversion of reduced rec MRP14 (peak 2) to oxidized rec MRP14 containing a disulfide bond between Cys_{79} and Cys_{90} (peak 1). Proteins were dialyzed in ammonium acetate, pH ~ 7, for (A) T = 0 h, (B) T = 24 h, (C) T = 48 h, and (D) T = 96 h. The identity of peak 3 is unknown (see text for details).

TABLE 2

Summary of the Expression, Digestion, and Isolation of rec MRP14

Step	Protein	Total protein (mg)	Yield (%)
Culture/liter	GST–MRP14	4.7	—
Thrombin (1 h)	MRP14	1.5	95
Thrombin (14 h)	$MRP14_{1-102}$	1.4	95
Oxidation (96 h)	$MRP14_{oxidized}$	0.8	55

The three Cys residues in mMRP14 can potentially form three distinct intramolecular disulfide-bonded isomeric forms, Cys_{79}-Cys_{90}, Cys_{90}-Cys_{110}, and Cys_{79}-Cys_{110}. Endoprotease Glu C digest products of native mMRP14 were characterized by ESI/MS and Edman sequencing after separating peptides by C18 RP-HPLC (13). The peptide $MRP14_{79-92}$ was isolated and contained an intramolecular disulfide bond (13). The intramolecular disulfide bond produced after oxidative refolding of rec mMRP14 was located by comparing C18 RP-HPLC traces of Glu C digest products of native mMRP14, rec mMRP14, and oxidized rec mMRP14. Apart from slight differences in relative intensities, essentially the same chromatographic profiles were obtained for each protein and a peptide corresponding to $MRP14_{79-92}$ (rt = 18.3 min) was isolated from each digest. The mass of these peptides was 1640 ±2 Da and corresponded to the predicted mass of $MRP14_{79-92}$ (1641.1 Da reduced or 1639.1 Da oxidized). No peptides corresponding to predicted fragments with disulfide bonds to Cys_{110} were identified, supporting disulfide bond formation between the same Cys residues in native mMRP14 and oxidized rec mMRP14. A peptide corresponding to $MRP14_{79-92}$ (ESI mass 1640 Da) was also isolated from Glu C digestion of the iodoacetamide derivative of oxidized rec mMRP14, further supporting the location of the disulfide bond.

Zinc Binding

S100 proteins commonly bind Ca^{2+} and Zn^{2+} (30,31). S100β binds eight Zn^{2+} ions, four with high affinity and four with low affinity, probably via His residues, but the precise binding sites are uncharacterized (30). Ca^{2+} antagonizes low-affinity Zn^{2+} binding due to a conformational change (30). These properties may explain the multiple interactions and different extracellular and intracellular targets of S100β in the brain (5). Both human and murine MRP14 bind substantially more Zn^{2+} than either S100A8 (CP10) or S100β (13).

Equal quantities of native mMRP14, rec mMRP14, and rec $MRP14_{1-102}$ were separated on SDS–PAGE, blotted onto PVDF, and incubated with radioactive $^{65}ZnCl_2$. Native mMRP14, containing the disulfide bond between Cys_{79} and Cys_{90}, and reduced rec MRP14 bound approximately equal (±5%) amounts of $^{65}Zn^{2+}$ (Fig. 4, lanes 1 and 2, respectively) indicating that the Cys residues are not involved in Zn^{2+} binding, whereas binding was reduced by 80% (±5%) with similar quantities of rec $MRP14_{1-102}$ (Fig. 4, lane 3). All $^{65}Zn^{2+}$ binding was inhibited by excess $ZnCl_2$ (not shown), although excess $CaCl_2$ did not inhibit Zn^{2+} binding indicating that the binding site was distinct and independent from the two Ca^{2+} binding domains (Fig. 4, lanes 4, 5, and 6). Minor bands (<5%) at approximately 28,000 were attributed to disulfide-bonded dimeric forms of MRP14 (Fig. 4). Overlay experiments suggest that full-length mMRP14 has substantially higher affinity for Zn^{2+} than $MRP14_{1-102}$ (Fig. 4). Analysis of the amino acid sequence of mMRP14 does not indicate putative "zinc fingers" (32) or clusters (4–12) of acidic amino acids (33) which potentially may bind Zn^{2+}, but the protein contains seven His and an unmodified Cys, some of which may adopt an appropriate conformation for binding (34). Zinc binding has also been proposed to occur at His residues in the sequence HEXXH in hMRP14 and may be implicated in its ability to inhibit antimicrobial growth (11). mMRP14 has two such motifs, located at either side of the putative thrombin cleavage site, and both could potentially bind Zn^{2+}. The C-terminal peptide generated by thrombin digestion has the sequence GHGHSHGKGCGK. The three His and one Cys residue could also potentially chelate Zn^{2+} independently or with other residues located within mMRP14. Additional experiments are required to define the precise Zn^{2+} binding sites of mMRP14, but experiments reported here confirm the involvement of the extreme C-terminal motif in Zn^{2+} chelation.

In conclusion, we have characterized two forms of rec mMRP14, the full-length protein and a truncated

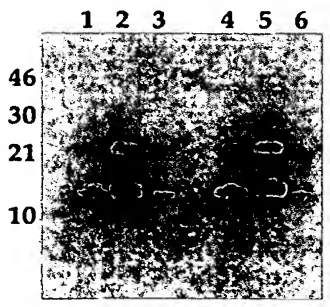


FIG. 4. $^{65}Zn^{2+}$ binding after transfer of proteins separated by SDS–PAGE to PVDF, native mMRP14 (lane 1), rec mMRP14 (lane 2), and rec $MRP14_{1-102}$ (lane 3) in TBS and binding to native mMRP14 (lane 4), rec mMRP14 (lane 5), and rec $MRP14_{1-102}$ (lane 6) in TBS + $CaCl_2$ (1 mM).

form, $MRP14_{1-102}$, produced by thrombin cleavage at an alternate site. Interestingly, MRP14 released by necrotic neutrophils at sites of inflammation (35) would be located in regions of active fibrin formation and the susceptibility of this site to thrombin cleavage may have functional relevance. Oxidative refolding of the recombinant protein readily occurred at pH ~ 7 in ammonium acetate buffer yielding a protein with intramolecular disulfide bonds equivalent to those found in the native protein. This will allow more detailed structural characterization of mMRP14 and more reliable functional analysis. Zn^{2+} binding by the truncated form of mMRP14 was greatly reduced, suggesting that residues in this domain are important in Zn^{2+} binding to mMRP14. The serendipitous identification of a thrombin cleavage product with reduced capacity to bind Zn^{2+} may have physiological relevance.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Health and Medical Research Council of Australia. Members of the Cytokine Research Unit are acknowledged for helpful discussions.

REFERENCES

1. Schäfer, B. W., Wicki, R., Engelkamp, D., Mattei, M. G., and Heizmann, C. W. (1995) Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21: Rationale for a new nomenclature of the S100 calcium-binding protein family. *Genomics* **25**, 638–643.
2. Hessian, P. A., Edgeworth, J., and Hogg, N. (1993) MRP-8 and MRP-14, two abundant Ca^{2+}-binding proteins of neutrophils and monocytes. *J. Leukocyte Biol.* **53**, 197–204.
3. Zimmer, D. B., Cornwall, E. H., Landar, A., and Song, W. (1995) The S100 protein family: History, function, and expression. *Brain Res. Bull.* **37**, 417–429.
4. Kligman, D., and Hilt, D. C. (1988) The S100 protein family. *TIBS* **13**, 437–443.
5. Fano, G., Biocca, S., Fulle, S., Mariggio, M. A., and Belia, S. (1995) The S-100—A protein family in search of a function. *Prog. Neurobiol.* **46**, 71–82.
6. Hessian, P. A., Wilkinson, L., and Hogg, N. (1995) The S100 family protein MRP-14 (S100A9) has homology with the contact domain of high molecular weight kininogen. *FEBS Lett.* **371**, 271–275.
7. Watt, K. W. K., Brightman, I. L., and Goetzl, E. J. (1983) Isolation of two polypeptides comprising the neutrophil-immobilizing factor of human leucocytes. *Immunology* **48**, 79–86.
8. Newton, R. A., and Hogg, N. (1998) The human S100 protein MRP-14 is a novel activator of B_2 integrin Mac-1 on neutrophils. *J. Immunol.* **160**, 1427–1435.
9. Johne, B., Fagerhol, M. K., Lyberg, T., Prydz, H., Brandtzæg, P., Naess-Anderson, C. F., and Dale, I. (1997) Functional and clinical aspects of the myelomonocytic protein calprotectin. *J. Clin. Pathol. Mol. Pathol.* **50**, 113–123.
10. Sohnle, P. G. (1997) Antimicrobial defense through competition for zinc. *Rev. Med. Microbiol.* **8**, 217–224.
11. Loomans, H. J., Hahn, B. L., Qi-Qin, L., Phadnis, S. H., and Sohhnle, P. G. (1998) Histidine-based zinc-binding sequences and the antimicrobial activity of calprotectin. *J. Infect. Dis.* **177**, 812–814.
12. Lackmann, M., Cornish, C. J., Simpson, R. J., Moritz, R. L., and Geczy, C. L. (1992) Purification and structural analysis of a murine chemotactic cytokine (CP-10) with sequence homology to S-100 proteins. *J. Biol. Chem.* **267**, 7499–7504.
13. Raftery, M. J., Harrison, C. A., Alewood, P., Jones, A., and Geczy, C. L. (1996) Isolation of the murine S100 protein MRP14 from activated spleen cells: Characterization of post translational modifications and zinc binding. *Biochem. J.* **316**, 285–293.
14. Iismaa, S. E., Hu, S., Kocher, M., Lackmann, M., Harrison, C. A., Thliveris, S., and Geczy, C. L. (1994) Recombinant and cellular expression of the murine chemotactic protein, CP-10. *DNA Cell Biol.* **13**, 183–192.
15. Raftery, M. J., Harrison, C. H., and Geczy, C. L. (1997) Characterization of a mutant recombinant S100 protein using electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **11**, 405–409.
16. Weir, M. P., and Sparks, J. (1987) Purification and renaturation of recombinant human interleukin-2. *Biochem. J.* **245**, 85–91.
17. Cleland, J. L. (1993) *in* "ASC Symposium Series 526," Vol. 526, Am. Chem. Soc., Washington, DC.
18. Han, B., Hall, F. L., and Mimni, M. E. (1997) Refolding of a recombinant collagen-targeted TGF-β2 fusion protein expressed in *Escherichia coli*. *Protein Express. Purif.* **11**, 169–178.
19. Scotto, C., Mély, Y., Ohshima, H., Garin, J., Cochet, C., Chambaz, E., and Baudier, J. (1998) Cysteine oxidation in the mitogenic S100B protein leads to changes in phosphorylation by catalytic CKII-α subunit. *J. Biol. Chem.* **273**, 3901–3908.
20. Schägger, H., and von Jagow, G. (1987) Tricine–sodium dodecyl sulphate–polyacrylamide electrophoresis for the separation of peptides in the range 1 to 100 kDa. *Anal. Biochem.* **166**, 368–379.
21. Lagasse, E., and Weissman, I. L. (1992) Mouse MRP8 and MRP14, two intracellular calcium-binding proteins associated with the development of the myeloid lineage. *Blood* **79**, 1907–1915.
22. Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusion proteins with glutathione-*S*-transferase. *Gene* **67**, 31–40.
23. Tuggle, C. K., and Fuchs, J. A. (1985) Glutathione reductase is not required for maintenance of reduced glutathione in *Escherichia coli* K-12. *J. Bacteriol.* **162**, 448–450.
24. Ashton, D. S., Beddell, C. R., Green, B. N., and Oliver, R. W. A. (1994) Rapid validation of molecular structures of biological samples by electrospray–mass spectrometry. *FEBS Lett.* **342**, 1–6.
25. Raftery, M. J., and Geczy, C. L. (1998) Identification of post translational modifications and cDNA sequencing errors in the rat S100 proteins MRP8 and 14 using electrospray ionization mass spectrometry. *Anal. Biochem.* **258**, 285–292.
26. Pedrocchi, M., Schafer, B. W., Durussel, I., Cox, J. A., and Heizmann, C. W. (1994) Purification and characterization of the recombinant human calcium-binding S100 proteins CAPL and CACY. *Biochemistry* **33**, 6732–6738.
27. Föhr, U. G., Heizmann, C. W., Engelkamp, D., Schäfer, B. W., and Cox, J. A. (1995) Purification and cation binding properties of the recombinant human S100 calcium-binding protein A3, an EF-hand motif protein with high affinity for zinc. *J. Biol. Chem.* **270**, 21056–21061.
28. Chang, J.-Y. (1985) Thrombin specificity: Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate. *Eur. J. Biochem.* **151**, 217–224.

29. Rainer, R., and Hauke, L. (1996) *In vitro* folding of inclusion body proteins. *FASEB J.* **10**, 49–56.
30. Baudier, J., Glasser, N., and Gerard, D. (1986) Ions binding to S100 proteins. *J. Biol. Chem.* **261**, 8192–8203.
31. Filipek, A., Heizmann, C. W., and Kuźnicki, J. (1990) Calcyclin is a calcium and zinc binding protein. *FEBS Lett.* **264**, 263–266.
32. Evans, R. M., and Hollenberg, S. M. (1988) Zinc fingers: Gilt by association. *Cell* **52**, 1–3.
33. Brand, I. A., Heinickel, A., Kratzin, H., and Söling, D.-H. (1988) Properties of a 19-kDa Zn^{2+}-binding protein and sequence of the Zn^{2+}-binding domains. *Eur. J. Biochem.* **177**, 561–568.
34. Vallee, B. L., and Auld, D. S. (1990) Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* **29**, 5647–5659.
35. Kocher, M., Kenny, P. A., Farram, E., Majid, K., Finlay-Jones, J. J., and Geczy, C. L. (1996) Functional chemotactic factor CP-10 and MRP-14 are abundant in murine abscesses *Infect. Immunol.* **64**, 1342–1350.

JOURNAL OF BACTERIOLOGY, Feb. 1999, p. 934–940
0021-9193/99/$04.00+0

Vol. 181, No. 3

The Amino-Terminal 100 Residues of the Nitrogen Assimilation Control Protein (NAC) Encode All Known Properties of NAC from *Klebsiella aerogenes* and *Escherichia coli*

WILSON B. MUSE AND ROBERT A. BENDER*

Department of Biology, The University of Michigan, Ann Arbor, Michigan 48109-1048

Received 14 August 1998/Accepted 21 November 1998

The nitrogen assimilation control protein (NAC) from *Klebsiella aerogenes* or *Escherichia coli* (NAC_K or NAC_E, respectively) is a transcriptional regulator that is both necessary and sufficient to activate transcription of the histidine utilization (*hut*) operon and to repress transcription of the glutamate dehydrogenase (*gdh*) operon in *K. aerogenes*. Truncated NAC polypeptides, generated by the introduction of stop codons within the *nac* open reading frame, were tested for the ability to activate *hut* and repress *gdh* in vivo. Most of the NAC_K and NAC_E fragments with 100 or more amino acids (wild-type NAC_K and NAC_E both have 305 amino acids) were functional in activating *hut* and repressing *gdh* expression in vivo. Full-length NAC_K and NAC_E were isolated as chimeric proteins with the maltose-binding protein (MBP). NAC_K and NAC_E released from such chimeras were able to activate *hut* transcription in a purified system in vitro, as were NAC_K129 and NAC_E100 (a NAC_K fragment of 129 amino acids and a NAC_E fragment of 100 amino acids) released from comparable chimeras. A set of NAC_E and NAC_K fragments carrying nickel-binding histidine tags (his_6) at their C termini were also generated. All such constructs derived from NAC_E were insoluble, as was NAC_E itself. Of the his_6-tagged constructs derived from NAC_K, NAC_K100 was inactive, but NAC_K120 was active. Several NAC fragments were tested for dimerization. NAC_K120-his_6 and NAC_K100-his_6 were dimers in solution. MBP-NAC_K and MBP-NAC_K129 were monomers in solution but dimerized when the MBP was released by cleavage with factor Xa. MBP-NAC_E was readily cleaved by factor Xa, but the resulting NAC_E was also degraded by the protease. However, MBP-NAC_E-his6 was completely resistant to cleavage by factor Xa, suggesting an interaction between the C and N termini of this protein.

The nitrogen regulatory (Ntr) system of enteric bacteria allows *Klebsiella aerogenes*, *Escherichia coli*, and other enteric bacteria to respond rapidly and effectively to changes in the quality of the nitrogen source provided (13). Under conditions of nitrogen limitation, a complex cascade of regulatory events involving uridylylation and phosphorylation reactions leads to the accumulation of the phosphorylated (active) form of the transcriptional activator NtrC. NtrC-phosphate can activate the σ^{54}-dependent expression of a variety of genes involved in utilization of organic and inorganic nitrogen sources. NtrC-phosphate also activates transcription of the *nac* gene in *K. aerogenes* and *E. coli* (6).

The nitrogen assimilation control protein (NAC) is a regulatory protein responsible for activating the transcription of operons such as *hutUH*, *putP*, and *ureDABCEFG*, whose products supply the cell with ammonia or glutamate from histidine, proline, and urea, respectively (2). NAC is also responsible for repressing transcription of the *gdhA* and *gltBD* operons, whose products are involved in assimilating ammonia under conditions of nitrogen excess or limitation (2). NAC is also responsible for down-regulating its own transcription (3, 7). The expression of the *nac* gene is entirely dependent on the Ntr system such that under conditions of nitrogen-limited growth, the Ntr system is active and *nac* is actively transcribed by RNA polymerase charged with the minor sigma factor σ^{54} (6, 12). There are no known coeffectors involved in the regulation of transcription by NAC (17), so if *nac* is expressed, then NAC is produced and activates transcription of *hut*, *put*, and *ure* by RNA polymerase charged with the major sigma factor σ^{70}. Thus, NAC serves as a coupling factor between the σ^{54}-dependent Ntr system and σ^{70}-dependent genes like *hut* (2).

NAC is a member of the large family of LysR-type transcriptional regulators (LTTRs; 15, 16). This group of over 50 proteins are identified on the basis of amino acid sequence similarity. Most of the members of this family have between 300 and 350 amino acids and show a surprisingly high degree of sequence similarity. For example, NAC from *K. aerogenes* shows about 40% sequence identity with OxyR from *E. coli* (16). Within the LTTRs, the DNA-binding domain occurs near the N terminus of the protein, where a predicted helix-turn-helix motif (approximately amino acids 20 to 40 in NAC) is located (15). In several LTTRs, the binding site for a required coeffector has been identified. These binding sites are generally in the C-terminal third of the polypeptide, suggesting that the DNA-binding and coeffector-binding domains are physically separated (15).

Very little structural information about LTTRs is available. A fragment of CysB containing the C-terminal 233 amino acids has been crystallized, and a structure was derived from X-ray diffraction data (18). However, this fragment lacks the N-terminal 87 amino acids and thus the DNA-binding domain. As a result, this structure says little about how the DNA-protein interactions occur, are altered by the presence of inducer, or result in the activation of transcription.

Comparison of NAC from *E. coli* (NAC_E) with NAC from *K. aerogenes* (NAC_K) revealed a surprising lack of sequence similarity (14). NAC_E was only about 75% identical to NAC_K, in contrast to other *E. coli* and *K. aerogenes* regulatory proteins, which are >95% identical. Most of the sequence divergence

* Corresponding author. Mailing address: Department of Biology, The University of Michigan, 830 N. University, Ann Arbor, MI 48109-1048. Phone: (734) 936-2530. Fax: (734) 647-0884. E-mail: rbender@umich.edu.

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Description	Source or reference
KC2668	*hutC515* Δ[*bla*]-2 *nac*$^+$	10
KC2725	*hutC515* Δ[*bla*]-2 *nac-203*::Tn*5-131*	10
KC2972	KC2941/pCB511	Transformation
KC3220	KC2725/pCB594	Transformation
KC3555	KC2725/pCB771	Transformation
KC3568	KC2725/pCB774	Transformation
KC3572	KC2725/pCB778	Transformation
KC3578	KC2725/pCB781	Transformation
KC3580	KC2725/pCB782	Transformation
KC3591	KC2725/pCB788	Transformation
KC3615	KC2725/pCB800/pCB648	Transformation
KC3973	KC2725/pCB941	Transformation
KC3977	KC2725/pCB529	Transformation
KC3981	KC2725/pCB606	Transformation
KC4510	KC2725/pCB547	Transformation
Plasmids		
pCB511	pGB2; wild-type nac_E[a]	This laboratory
pCB529	pGB2; *nac-10* allele of *E. coli*	This laboratory
pCB547	pGB2; *nac-28* allele of *E. coli*	This laboratory
pCB594	pMalC2; MBP-NAC_E fusion protein	This laboratory
pCB606	pMalC2; MBP-NAC_K fusion protein	This laboratory
pCB771	pGD103; NAC_E-his_6	This laboratory
pCB774	pGD103; wild-type NAC_E	This laboratory
pCB778	pGD103; wild-type NAC_K	This laboratory
pCB781	pGD103; NAC_K-his_6	This laboratory
pCB782	pGD103; NAC_K-(DFGRSGHTDSL)	This laboratory
pCB788	pGD103; NAC_K100-his_6	This laboratory
pCB800	pGD103; NAC_K120-his_6	This laboratory
pCB941	pGD103; NAC_E100-his_6	This laboratory
pGB2	Low-copy cloning vector; insert expressed from its own promoter	4
pGD103	Low-copy vector expressing inserted material from *lacP*	5
pMalC2	MBP fusion vector expressing inserted material from *tac* promoter	New England BioLabs

[a] Vector; cloned insert or product.

between NAC_E and NAC_K occurs in the C-terminal two-thirds of the protein, consistent with the lack of a need to conserve a coeffector-binding site in the C-terminal domain. This raised the question of whether the C-terminal domain has any role at all, leading us to seek the smallest N-terminal fragment that would retain the biological activities associated with NAC.

Two *nac* mutations also suggested that a C-terminal portion of NAC might be dispensable, further encouraging us to search for a small active fragment of NAC. The *nac-112* allele of *K. aerogenes* is an insertion of Mud*lac,amp* at an unknown site within nac_K, which retains considerable NAC activity (12). The *nac-10* allele of *E. coli*, an insertion of a stop codon and a drug resistance cartridge in the center of the nac_E gene, was constructed as an intermediate in constructing the *nac-28* null allele (14). Strains carrying *nac-10* also retained considerable NAC activity. Thus, we attempted to determine whether the C-terminal domain of NAC from *E. coli* and *K. aerogenes* plays an essential role and to define the smallest N-terminal fragment of NAC that would retain the ability to function as a transcriptional regulator.

MATERIALS AND METHODS

Strains and plasmids. All of the *K. aerogenes* strains used in this study were derived from W70 and are listed in Table 1.

Generation of C-terminal deletions of NAC. Unidirectional exonuclease III deletions were performed to make 3′ deletions in the *E. coli nac* gene by using the protocol described by Henikoff (9) with minor modifications. Plasmid pCB574 contains the *E. coli nac* gene inserted into the *Eco*RI site of plasmid pGB2 (4) such that the 3′ end of the gene is proximal to the *Pst*I site present in the multiple cloning site. Deletions were initiated by cutting with *Sal*I and *Pst*I to allow unidirectional deletions into *nac*. After digestion with exonuclease III and mung bean nuclease, the deleted DNA fragments were resuspended in 20 μl of ligation buffer (250 mM Tris-HCl [pH 7.6], 50 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 mM ATP) and 100 pmol of a dephosphorylated *Nhe*I linker (containing stop codons in three reading frames) to which T4 DNA ligase was added. The ligation mixture was used to transform *E. coli* DH5α to streptomycin (50 μg/ml) and spectinomycin (100 μg/ml) resistance. Deletions of interest were sequenced by using the pUC reverse primer to determine the precise endpoint of each deletion. These were brought under the control of the *lac* promoter by PCR amplification by using the primers CGAATTCAAACTGGAGACTCATATGAAC (forward) and AGGATCCTCACACAGGAAACAGCTATGAC (reverse) and insertion into the *Eco*RI and *Bam*HI sites of pGD103 (5). Specific truncations in the *K. aerogenes nac* gene were generated by PCR using a forward primer (CTGGAATTCCTTACAGGAGGCA) containing an *Eco*RI site and reverse primers complementary to the last 18 bases prior to the desired truncation point followed by an amber stop codon and GGATCC (*Bam*HI). PCR products were introduced into the *Eco*RI and *Bam*HI sites of plasmid pGD103, and the sequences were verified by dideoxy sequencing.

Purification of MBP-NAC fusions. The *nac* genes from *E. coli* and *K. aerogenes*, along with the C-terminal truncated derivatives of each, were inserted into vector pMalC2 (New England Biolabs) as a blunt-ended 5′ end of the *nac* gene beginning at the initiating methionine ligated to the blunt-end *Xmn*I site within the vector. The 3′ ends were ligated into either the *Eco*RI or *Bam*HI sites within the multiple cloning site. The initial construct was prepared by PCR and sequenced to verify that errors had not been introduced by PCR amplification. Later derivatives were made by replacing sequences within this clone by using internal sites (*Xmn*I, *Ssp*I, and *Bgl*II) in the *nac* gene rather than reamplifying for each construct. Insertion of a blunt-ended DNA fragment into the *Xmn*I site of the vector results in fusion of the NAC protein to the 42-kDa maltose-binding protein (MBP). The junction between the two proteins contains the recognition site for factor Xa protease (Ile Glu Gly Arg) positioned such that cleavage with this enzyme would result in separation of the intact NAC protein from the fusion.

Purification of the fusion protein was initiated by inducing mid-log-phase cultures (3×10^8 CFU/ml) with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The induced cultures (500 ml in 2-liter flasks) were shifted from 37 to 28°C and grown for an additional 2 h. Cells were chilled on ice and harvested by centrifugation. The pellets were washed twice in column buffer (100 mM Tris-HCl [pH 7.5], 250 mM NaCl, 1 mM dithiothreitol, 1 mM 2-mercaptoethanol) and resuspended in a final volume of 15 ml. Cell disruption was accomplished by passage through a French pressure cell (twice at 12,000 lb/in^2). The resulting lysate was clarified by centrifugation for 30 min at 30,000 × *g* at 4°C. The supernatant solution containing the fusion protein was increased in volume to 30 ml with column buffer containing 50 mM freshly prepared phenylmethylsulfonyl fluoride. The crude preparation was applied to a column (2.5 cm in diameter) containing a 10-ml bed volume of amylose-agarose resin (New England Biolabs no. 800-21L) at a flow rate of 1 ml/min. Bound fusion protein was washed with 300 ml of column buffer at a 2-ml/min flow rate and eluted with column buffer supplemented with 1 M maltose at a rate of 1.5 ml/min with 3-ml fractions collected over 50 ml. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to determine which fractions contained MBP-NAC. The MBP-NAC fusion generally eluted over a wide peak from fractions 3 to 7. Fractions were pooled if SDS-PAGE indicated sufficient purity. Protein concentrations were determined by Bio-Rad protein assay and generally indicated values between 0.7 and 2.5 mg/ml. Diafiltration was performed by using either an Amicon 10- or 25-ml stir cell apparatus with PM30 filters to concentrate samples. Fusions were stored at −20°C following the addition of an equal volume of glycerol. If the fusion was to be cleaved or used immediately, it was kept on ice at all times. Diafiltration was also used to change the buffer in which the MBP-NAC was suspended to buffer 4 (8), which contains 100 mM NaH_2PO_4 (pH 7.0), 250 mM NaCl, 2.5 mM $MgCl_2$, and 1 mM 2-mercaptoethanol. The NAC protein was not transcriptionally active as a fusion protein with MBP. Therefore, to retrieve active NAC from these fusions for in vitro transcription experiments, it was necessary to cleave NAC from the fusion by using factor Xa protease. This was accomplished by adding 1 μg of protease to 0.5 mg of the fusion protein in a 100-ml volume and then incubating it for 16 h at 22°C. Following cleavage, freshly prepared phenylmethylsulfonyl fluoride was added to 20 mM. The cleaved NAC fusion protein had to be used immediately because it degraded quickly over time.

Purification of his_6-tagged NAC. To overcome the problems associated with MBP-NAC fusions, various six-histidine (his_6)-tagged constructs were built. NAC N-terminal constructs were made by cloning *nac* genes into the *Eco*RI/*Hin*dIII sites of pET28 (Novagen). The N-terminal fusions added a leader to the NAC proteins consisting of 56 amino acids which include a stretch of six histidines in a row to facilitate nickel-nitrilotriacetic acid affinity purification (Qiagen). Constructs in which carboxy-terminal his_6 tags were added had the DNA sequence $(CAT)_6$ followed by an amber stop codon designed into the

primers used to amplify the *nac* gene. The C-terminal his_6-tagged genes were inserted into a derivative of the pET28 vector in which the material between the *Xho*I and *Eco*RI sites had been removed. Histidine-tagged NAC proteins were expressed as previously described for the MBP fusions and prepared for affinity purification in a similar fashion, except that the pH of the buffer was raised to 7.8 to facilitate nickel interaction. Clarified S30 French pressate was added to a column with an internal diameter of 1.25 cm containing a 10-ml bed volume of nickel-nitrilotriacetic acid resin (Qiagen). NAC was allowed to bind the column at a flow rate of 2 ml/min. The column was washed with 150 ml of column buffer supplemented with 5 mM imidazole to eliminate nonspecific binding. NAC was eluted with elution buffer (300 mM imidazole [pH 7.5], 250 NaCl, 1 mM $MgCl_2$, 1 mM 2-mercaptoethanol). Fractions of 3 ml were collected and screened by Bio-Rad protein assay. NAC generally eluted in a tight peak in fraction 3. Purity was monitored by SDS-PAGE.

Gel filtration analysis of purified proteins. Gel filtration chromatography was used to determine the size and oligomeric state of purified NAC proteins. Size determinations were made by one of two methods: fast protein liquid chromatography (FPLC)-mediated gel filtration or gravity-fed chromatography. For determinations by FPLC, a Pharmacia LCC-501 FPLC apparatus was used in conjunction with a Superdex 75 HR 10/30 prepacked column. The column was equilibrated with buffer 4 (described above) and loaded with 30 μl (75 μg) of NAC. Separation was allowed to continue at a flow rate of 1 ml/min for 40 min. Detection of eluting protein was done by monitoring A_{280}. Standardization of the column was done with gel filtration protein standards purchased from Sigma Chemical Co. resuspended in buffer 4. Generally, dimeric NAC (66-kDa dimer) eluted at a 15-ml volume at the same fraction as bovine serum albumin (66 kDa). Gravity flow gel filtration was performed by using an SR 25/100 column (25-mm internal diameter) packed with a 150-ml bed volume of Sephacryl S-100. Standardization of this column was done with the same buffers, globular weight standards, and concentrations as the FPLC, but the loading volume was increased to 1 ml. The flow rate was adjusted to 0.5 ml/min, and 150 ml was allowed to flow. Generally, dimeric NAC (66 kDa) eluted at a volume of 16 ml from this column.

In vitro transcription assays. In vitro transcriptions were carried out essentially as described previously (8). The template used was plasmid pCB695, which carries the *K. aerogenes hutUH* promoter inserted into vector pTE103, similar to pAM1202, which was described previously (8). Transcription mixtures contained 0.16 pmol of supercoiled plasmid as the template and 2 to 6 pmol of RNA polymerase in a total volume of 25 μl.

Enzyme assays. Histidase and glutamate dehydrogenase (GDH) activities were measured as described previously (12). Specific activities are reported as nanomoles of urocanate formed (histidase) or $NADPH_2$ oxidized (GDH) per minute per milligram of cell protein. Cell protein was determined by the method of Lowry et al. (11).

Gel mobility shift assays. Gel mobility shift assays were performed by using purified NAC_E or NAC_K or fragments thereof. The ability of NAC to bind DNA was determined by using the 330-bp *Eco*RI-to-*Hin*dIII fragment of the *ureD* promoter, which contains a strong NAC-binding site. Binding reaction mixtures consisted of 1 μl of DNA (0.05 pmol), 1 μl of poly[d(I·C) (50 ng/ml), 4 μl of deionized distilled H_2O, and 1 μl of a NAC dilution (0.35 to 1.7 pmol) in 50 mM NaH_2PO_4 (pH 7)–125 mM NaCl–0.5 mM $MgCl_2$–0.1 mM β-mercaptoethanol–50% glycerol, 1-mg/ml bovine serum albumin. NAC was incubated with the DNA for 30 min before addition of 1.5 μl of loading buffer (40 mM Tris [pH 8.4], 4 mM EDTA, 0.2% bromthymol blue, 0.2% xylene cyanol, 15% Ficoll) and applied to a 2% Tris-borate-EDTA agarose gel. Electrophoresis was carried out at room temperature at 15 V/cm for approximately 1 h. Gels were stained with ethidium bromide, and the fragments were visualized by UV fluorescence.

RESULTS

Residual NAC activity in *nac-10* mutants. The *nac-10* allele of the *E. coli nac* gene has a stop codon and a kanamycin resistance gene inserted into the middle of nac_E after codon 165. The resulting NAC polypeptide has four amino acids encoded by the polylinker fused to the 165th amino acid of NAC; the C-terminal 140 amino acids of NAC are not present. The *nac* gene of *E. coli* was replaced with *nac-10* as described elsewhere (14), and the replacement was confirmed by Southern blotting (data not shown). To test for NAC activity, we measured the ability of a strain carrying *nac-10* to activate the NAC-dependent expression of the *hut* operon in response to nitrogen-limited growth. The *nac-10* mutant showed about half as much histidase as the wild type under NAC-dependent conditions (data not shown), suggesting that the NAC protein encoded by *nac-10* retained considerable activity. However, NAC-dependent activation of histidase expression in *E. coli* is weak, making it dangerous to draw strong conclusions based on small differences.

TABLE 2. Residual activity of *nac-10* mutation[a]

Strain	Chromosomal allele	Plasmid allele	Presence of N in medium	Activity[b] of:	
				Histidase	GDH
KC2668	*nac*$^+$	None	+	28	410
			−	389	65
KC2725	*nac-203*	None	+	31	413
			−	54	436
KC2972	*nac-203*	*nac*$^+$	+	31	292
			−	460	41
KC3977	*nac-203*	*nac-10*	+	35	470
			−	265	141
KC4510	*nac-203*	*nac-28*	+	33	397
			−	59	425

[a] Strains were grown to mid-log phase in glucose minimal medium with 0.2% (wt/vol) glutamine as the sole nitrogen source (−) or with a combination of 0.2% glutamine and 0.2% ammonium sulfate as the nitrogen source (+) and assayed for histidase and GDH (12).

[b] Specific activity is reported as nanomoles per minute per milligram of cell protein.

NAC-dependent activation of histidase expression is stronger in *K. aerogenes*, and repression of glutamate dehydrogenase formation by NAC serves as another measure of NAC activity. Therefore, we studied the effect of the *nac-10* allele in *K. aerogenes*. The *nac-10* allele was cloned in low-copy expression vector pGD103 and tested for the ability to complement a *nac-203* mutation in *K. aerogenes*. The wild-type *nac* gene from *E. coli* was able to complement nac_K*-203* fully, showing 15-fold activation of histidase formation and 7-fold repression of GDH expression in response to nitrogen starvation. This is similar to the effect seen in wild-type (*nac*$^+$) *K. aerogenes* (Table 2, cf. KC2668 and KC2972). In contrast, a known *E. coli* null allele, nac_E*-28*, failed to complement nac_K*-203* with strain KC4510 (*nac-28/nac-203*), showing no more activation of histidase or repression of GDH expression than the *nac-203* mutant, KC2725. Strain KC3977 (*nac-10/nac-203*) showed about sevenfold activation of histidase formation and threefold repression of GDH formation in response to nitrogen starvation. Thus, the *nac-10* mutation, with an insertion of a stop codon and a kanamycin resistance cassette after amino acid 165, retained about half as much ability to activate and repress transcription as wild-type *nac*$^+$ from *E. coli*, despite the absence of the entire C-terminal half of NAC.

This residual activity is particularly surprising compared to the nac_K*-203* mutation of *K. aerogenes*. *nac-203* is an insertion of Tn*5* (or Tn*5-131*) after the 199th codon of nac_K and has one of the tightest Nac$^-$ phenotypes of all of the characterized *K. aerogenes nac* mutants (12). Therefore, we began a systematic search for the smallest NAC fragment with biological activity.

The N-terminal 100 amino acids of NAC are sufficient to activate and repress transcription. A collection of mutants with deletions of portions of the C-terminal part of the *K. aerogenes* and *E. coli nac* genes (nac_K and nac_E, respectively) was generated by exonuclease III digestion as described in Materials and Methods. These truncated *nac* genes were cloned into pGD103, where *nac* expression was under the control of the *lacZ* promoter. *K. aerogenes nac-203* mutant KC2725 was transformed with the resulting plasmids, and histidase and glutamate dehydrogenase activities were deter-

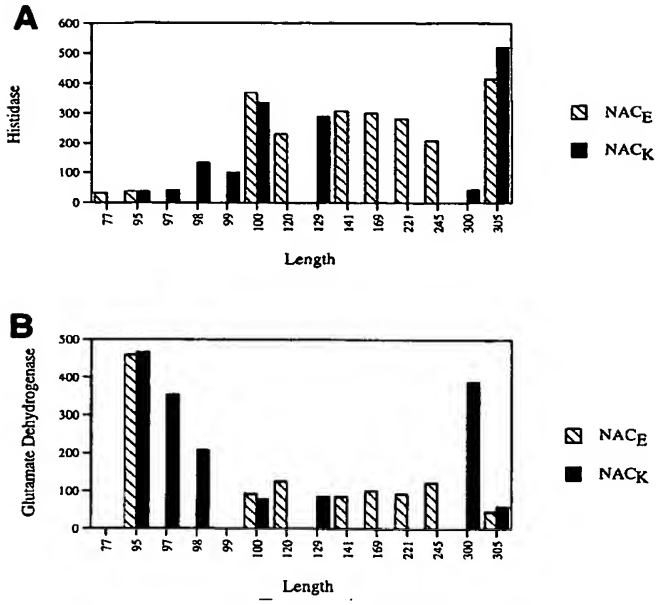


FIG. 1. *hut* activation (A) and *gdh* repression (B) by N-terminal fragments of NAC. *K. aerogenes* KC2725 (*nac-203*::Tn*5-131*) was transformed with derivatives of pGD103 containing an insert that coded for either full-length NAC or truncated versions of NAC. These transformants were assayed for NAC-dependent activation of histidase formation and repression of GDH formation. The values on the *x* axis represent the number of amino acids remaining in the NAC fragment. Thus, 305 represents the wild type and 100 represents the N-terminal 100 amino acids. Histidase and GDH values are reported as specific activities.

mined as a measure of NAC-dependent activation and repression. For convenience, the truncated proteins will be named by the number of the last remaining amino acid. Thus, wild-type *E. coli* NAC is NAC_E305 and a *K. aerogenes* NAC lacking the last five amino acids is NAC_K300.

Most of the truncated genes that retained 100 or more codons at the amino terminus were able to complement the nac_K-*203* allele for activation of histidase formation and repression of GDH formation (Fig. 1). The magnitude of the effect varied from about 50 to 90% as much activation or repression as the wild type, and the mutants retaining only 100 amino acids of NAC (NAC_K100 and NAC_E100) were among the most active of the truncations in vivo. This result was as true of the *E. coli nac* gene as of the *K. aerogenes nac* gene and appears to be a general property of NAC rather than a species-specific anomaly.

The single exception in this set was NAC_K300, which lacks only the five C-terminal amino acids of NAC_K and was totally inactive in complementation. Western blot assays of extracts prepared from cells carrying wild-type NAC_K or NAC_K300 were carried out by using anti-NAC serum. The wild type was readily seen, but NAC_K300 was not (data not shown), suggesting that the NAC_K300 product is unstable and degraded rapidly.

Although NAC_K100 has nearly full activity in vivo, NAC_K98 complements nac_K-*203* less well, and truncations with 77, 95, or 97 amino acids remaining failed to complement nac_K-*203*. We have not determined whether the failure of these very short NAC truncation products resulted from rapid degradation or an inactive protein. However, it is clear that a mutant *nac* gene that encodes only the first 100 of the 305 amino acids of NAC is functional in activation and repression of transcription, at least in vivo.

Activity of N-terminal fragments of NAC purified using MBP tags. Fragments of NAC containing only the N-terminal 100, 120, or 129 amino acids were nearly as active as wild-type NAC in vivo. In order to prove that these fragments were directly responsible for the activity, we attempted to demonstrate activation of transcription in vitro by using purified components. To purify these fragments, we fused the MBP to the N terminus via a peptide linker that could be cleaved with factor Xa protease, leaving an N-terminal methionine, as is seen in wild-type NAC. MBP-tagged variants of wild-type NAC (NAC_K305 and NAC_E305) and two N-terminal fragments (NAC_K129 and NAC_E100) were constructed. All of these fusion proteins were inactive in vivo.

NAC fragments released from MBP-NAC fusions were unstable in the presence of factor Xa (or thrombin when a thrombin-sensitive linker was used); therefore, the factor Xa-mediated cleavage products of MBP-NAC were added immediately to an in vitro transcription reaction mixture without further purification. As expected, cleaved MBP-NAC_K305 was able to activate *hut* transcription (Fig. 2A). Cleaved MBP-NAC_E305 was also able to activate *hut* transcription, consistent with the observation that the nac_E gene was able to complement a nac_K mutant in vivo. This also provided the first direct evidence that NAC_E was active in vitro as well as in vivo, since we had been unable purify NAC_E because of its insolubility under all of the conditions tested. Cleaved MBP-NAC_K129 and MBP-NAC_E100 were also able to activate *hut* transcription almost as well as cleaved MBP-NAC_K305 (Fig. 2B). Thus, NAC_K129 and NAC_E100 were able to activate *hut* operon expression in vitro as well as in vivo.

The instability of the NAC fragments in the presence of factor Xa precluded any attempt to quantify NAC in these reaction mixtures, as well as any attempt to isolate large quantities of NAC fragments. Thus, we attempted to characterize NAC fragments with a nickel-binding his_6 tag at the C terminus.

Activity of NAC fragments with his_6 tags. Full-length NAC_E305 and NAC_K305 with the nickel-binding his_6 tag at the C terminus were active in vivo. In fact, the activation of histidase and repression of GDH formation were even stronger with NAC305-his_6 that with the corresponding unmodified NAC305 (Table 3). This increased activation and repression appears not to be specific to the nature of the C-terminal tag, since a cloning artifact that fused an 11-amino-acid peptide (DFGRSGHTDSL) to the C terminus of NAC_K305 also resulted in more activation and repression than NAC_K305 (Table 3). Although the modified *E. coli* NAC (NAC_E305-his_6) was active in vivo, the modification did not prove useful as a means of purifying NAC_E, since NAC_E305-his_6 was insoluble in all of the buffers tested, just like NAC_E305.

Three N-terminal fragments with the his_6 tag at the C terminus were constructed as described in Materials and Methods: NAC_E100-his_6, NAC_K100-his_6, and NAC_K120-his_6. In contrast to their unmodified counterparts, both NAC_K100-his_6 and NAC_E100-his_6 were inactive in vivo (Table 3). Thus, we focused our attention on NAC_K120-his_6. NAC_K120-his_6 was purified by Ni affinity chromatography and shown to be able to activate *hut* operon transcription by purified RNA polymerase (data not shown). Unfortunately, we have had difficulty purifying NAC_K120-his_6 from strains carrying either high-copy or low-copy expression vectors on a reliable basis. We have no explanation for this difficulty, although it may represent either instability of the polypeptide or toxicity of the fragment.

In a final attempt to isolate an active form of NAC_E for

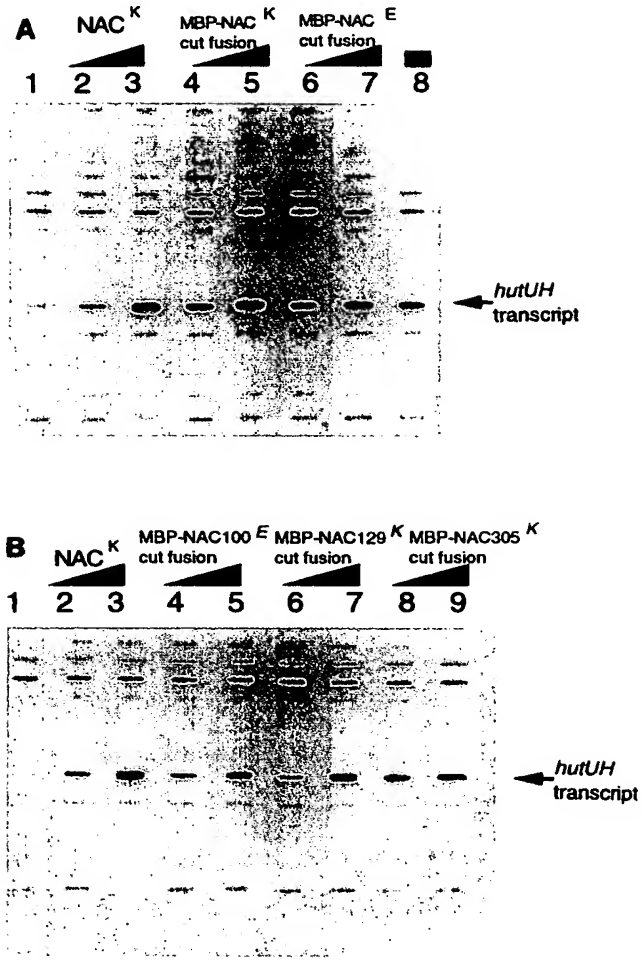


FIG. 2. Activation of *hutUH* transcription in vitro by NAC and N-terminal fragments of NAC. In vitro transcription was carried out by using purified components as described in Materials and Methods. The supercoiled template, pCB695, was similar to pAM1202 (8) and contained the *hutUH* promoter region cloned upstream of a strong transcriptional terminator such that transcription from this promoter would yield a transcript with the mobility on SDS-PAGE indicated at the right. All reactions contained RNA polymerase, nucleotides (including radioactive UTP), and buffers. NAC and NAC fragments were added as indicated. (A) Addition of full-length NAC. Lanes: 1, no NAC; 2 and 3, 0.04 and 0.4 μg of NAC (0.06 and 0.6 pmol of dimers) purified from *K. aerogenes* by salt precipitation (8); 4 and 5, 0.1 and 1 μg of MBP-NAC_K (1.3 and 13.3 pmol of MBP-NAC_K monomers) treated with factor Xa for 30 min immediately before addition to the transcription mixture; 6 and 7, 0.1 and 1 μg of MBP-NAC_E (1.3 and 13.3 pmol of monomers) treated with factor Xa for 30 min immediately before addition to the transcription mixture; 8, 3 μg of MBP-NAC_E (40 pmol of monomers) treated with factor Xa for 16 h before addition to the transcription mixture. (B) Addition of NAC fragments. Lanes: 1, no NAC; 2 and 3, 0.07 and 0.7 μg of NAC (1.1 and 10.5 pmol of dimers) purified from *K. aerogenes* by salt precipitation; 4 and 5, 0.1 and 1 μg of MBP-NAC_E100 (1.9 and 19 pmol of monomers) treated with factor Xa for 30 min immediately before addition to the transcription mixture; 6 and 7, 0.13 and 1.3 μg of MBP-NAC_K129 (1.9 and 19 pmol of monomers) treated with factor Xa for 30 min immediately before addition to the transcription mixture; 8 and 9, 0.14 and 1.4 μg of MBP-NAC_K305 (1.9 and 19 pmol of monomers) treated with factor Xa for 30 min immediately before addition to the transcription mixture.

comparison to NAC_K in vitro, we constructed a gene encoding NAC_E that was modified at the N terminus with MBP and at the C terminus with his_6 (MBP-NAC_E-his_6). We hoped to use the his_6 tag to separate the NAC_E fragment from factor Xa after cleavage of MBP-NAC_E-his_6. However, MBP-NAC_E-his_6 was completely resistant to cleavage by factor Xa, in marked contrast to MBP-NAC_E. Only the addition of strong denaturants allowed MBP to be removed, and this was accompanied by very rapid disappearance of NAC_E. No active NAC_E was recovered after attempts to renature the material. Apparently, the C-terminal his_6 tag had altered the factor Xa cleavage site in MBP-NAC_E-his_6 so as to make it inaccessible to cleavage by factor Xa, suggesting that the C terminus of the folded protein is close to the N terminus. Thus, the C-terminal his_6 tag would be able to prevent the protease from reaching the cleavage site at the N terminus of the NAC polypeptide.

TABLE 3. Activation of histidase formation and repression of GDH formation in vivo by modified forms of NAC

NAC source and strain[a]	Plasmid[b]	NAC product[c]	Sp act[d] of:	
			Histidase	GDH
E. coli				
KC2725	None	None	31	413
KC3568	pCB774	NAC_E	415	46
KC3210	pCB594	MBP-NAC_E	34	435
KC3555	pCB771	NAC_E-his_6	890	21
KC3973	pCB941	NAC_E100-his_6	42	354
K. aerogenes				
KC3572	pCB778	NAC_K	520	59
KC3981	pCB606	MBP-NAC_K	37	417
KC3578	pCB781	NAC_K-his_6	650	35
KC3580	pCB782	NAC_K-DFGRSGHDTSL	720	32
KC3615	pCB800	NAC_K120-his_6	295	ND[e]
KC3591	pCB788	NAC_K100-his_6	45	335

[a] All strains are derived from *K. aerogenes* KC2725 (*nac-203*::Tn*5-131*), grown in glucose minimal medium to mid-log phase, and assayed as previously described (12). IPTG (0.5 mM) was included for strain KC3981 to guarantee expression from the pMalC2 vector.

[b] All plasmids except pCB594 and PCB606 were derived from pGD103 as described in Materials and Methods. Plasmids pCB594 and pCB606 were derived from pMalC2.

[c] The code is described in detail in the text. Subscripts indicate derivation from *E. coli* (E) or *K. aerogenes* (K). MBP indicates MBP fused to the N terminus of NAC. his_6 indicates six histidine residues added to the C terminus. DFGRSGHTDSL indicates fusion of 11 amino acids to the C terminus of NAC (amino acids are indicated by the one-letter code). NAC100 and NAC120 indicate that only the first 100 or 120 amino acids of NAC were present.

[d] Specific activity is reported as nanomoles per minute per milligram of cell protein.

[e] ND, not determined.

Subunit structures. We next used the tagged NAC polypeptides to search for the dimerization site on NAC. These data (along with a summary of the in vitro activation data and DNA-binding data) are presented in Table 4. NAC_E and NAC_K with an N-terminal MBP tag eluted as monomers. If factor Xa was added, the elution profile was complex and included a peak of the size expected for MBP-NAC (uncleaved, monomer), MBP (monomer), NAC (dimer), and an unknown peak with an apparent size of 8.5 kDa (presumed to be a degradation product of the 33-kDa NAC monomer).

A smaller N-terminal modification was also tested by fusing a 56-amino-acid peptide that included a his_6 tag in its midst to the N terminus of NAC_K and NAC_E. Both of these fusions were inactive in vivo (data not shown). The N-terminally his_6-tagged NAC_E and NAC_K were purified by Ni affinity chromatography. The tagged NAC_K was found to be dimeric in solution even before cleavage (NAC_E was not tested). However, neither the NAC_K nor the NAC_E fusion protein was able to activate transcription in vitro (Table 4). Cleavage of the tagged NAC_E with thrombin removed the his_6 domain but left an

TABLE 4. In vitro properties of modified forms of NAC[a]

Construct	Apparent molecular size in solution (kDa)	Form[b]	Gel mobility shift[c]	Transcription activation[c]
NAC_E		I		
MBP-NAC_E	75 (66)	M (D)	Y (Y)	(Y)
56his_6-NAC_E			Y (Y)	N (N)
NAC_E-his_6		I		
NAC_E100-his_6		I		
MBP-NAC_E100				(Y)
NAC_K	66	D	Y	Y
MBP-NAC_K	70 (66)	M (D)	Y (Y)	N (Y)
56his_6-NAC_K	77	D	Y	
NAC_K-his_6		Y	Y	
MBP-NAC_K129	56 (28)	M (D)		(Y)
NAC_K120-his_6	27	D	Y	Y
NAC_K100-his_6	22	D	Y	

[a] NAC_E, NAC_K, and several modified forms thereof were purified as described in Materials and Methods. The apparent molecular size in solution was determined by gel filtration, and this size was used to infer whether the product was monomeric or dimeric in solution. The ability to bind DNA was determined in a gel mobility shift assay using the urease operon promoter (*ureDp*) as a target. The ability to activate transcription from the *hutUH* promoter (*hutUp*) was tested as described in the legend to Fig. 2. The values in parentheses were determined after cleavage with Factor Xa or thrombin.

[b] M, monomer; I, insoluble; D, dimer.

[c] Y, yes; N, no.

18-amino-acid peptide fused to the N terminus of NAC_E. This cleavage product was also unable to activate transcription in vitro (Table 4). In summary, fusions to the N terminus of NAC_E increase its solubility, but even small additions at the N terminus block its ability to activate transcription. Large additions (MBP) block the ability to dimerize in solution (Table 4).

Somewhat surprisingly, the C-terminally his_6-tagged NAC fragments of NAC_K (NAC_K100-his_6 and NAC_K120-his_6), as well as the product released from MBP-NAC_K129 by (factor Xa) cleavage, were all dimers in solution as measured by gel filtration. Thus, the N-terminal domain of 100 amino acids also appears to contain the signals required for dimer formation, as well as for activation of transcription.

DISCUSSION

Perhaps the most surprising finding in these experiments is that the N-terminal 100 amino acids of NAC contain the determinants for all of the known functions of NAC. NAC100 was able to activate histidase formation in vivo and *hutUp* transcription in vitro. NAC100 was able to repress GDH formation in vivo. NAC_K100-his_6 was able to bind a DNA fragment with a NAC-binding site, and NAC_K100-his_6 was present as a dimer in solution. In short, NAC_K100 (or NAC_K100-his_6) displayed all four known properties of NAC: activation of transcription, repression of transcription, site-specific DNA binding, and dimerization.

All of the truncations of NAC isolated in these experiments, except NAC_K300, showed considerable NAC activity. However, several tight *nac* mutants have been isolated before by using selection and screening. One of these, *nac-203*::Tn*5*, has been sequenced and is an insertion of Tn*5* after the 199th codon of nac_K. The data in Fig. 1 suggest that this mutation should leave NAC partially active, but *nac-203* is one of the tightest *nac* mutants in our collection (although even *nac-203* retains some activity, particularly on the *ure* operon promoter). Thus, it seems likely that NAC_K300 and NAC_K199 are inactive because of instability or misfolding rather than because of any lack of informative sequence.

The data presented here also provide some information about the function of the N and C termini of NAC. Additions to the N terminus block the ability of NAC to activate transcription, and small additions to the C terminus increase this activity, at least in vivo. The effect of the N-terminal additions appears to be intrinsic to the protein, since it is also seen in vitro with purified protein. Moreover, the ability to activate transcription can be recovered (in the case of MBP-NAC) by removing the N-terminal addition. The effect of the C-terminal additions may likewise be intrinsic to the protein, but it may also reflect an increased stability of the C-terminally modified NAC. We have been unable to raise a high-titer antibody with sufficient specificity for NAC to address this question directly, but we have previously suggested that wild-type NAC is rapidly turned over in vivo (14).

Two lines of evidence led us to suspect an interaction between the N and C termini of NAC and perhaps other LTTRs. First, MBP-NAC-his_6 was completely resistant to cleavage by factor Xa, while MBP-NAC was readily cleaved. Second, other work in this laboratory (1) has shown that replacement of the amino-terminal 20 amino acids of NAC with the corresponding amino acids from OxyR results in an inactive NAC unless the C-terminal domain of this NAC-OxyR chimera is also replaced with the C-terminal domain from OxyR.

The helix-turn-helix region thought to be the DNA-binding site is located very close to the N terminus of LTTRs (about amino acids 20 to 40 for NAC). The binding site for coeffectors of LTTRs has generally been found in the C-terminal domain, between amino acids 100 and 200 (15). Thus, an effect of the C-terminal domain on the N-terminal domain must exist. The specificity requirement in the OxyR-NAC chimeras (1) and the ability of the C-terminal his_6 tag to prevent cleavage of MBP from the N terminus of MBP-NAC-his_6 seem more consistent with a hypothesis of direct interaction than with transmission of a signal through the middle part of the protein. However, in the absence of experiments to test a direct interaction, such arguments remain speculative at best.

It was exciting to discover that the four known regulatory properties of NAC are encoded within a NAC fragment containing only the N-terminal 100 amino acids, NAC100. It was disappointing to discover that a fifth property of NAC, the insolubility of NAC_E, was also encoded within this fragment. Most of the amino acid differences between NAC_E and NAC_K lie in the C-terminal domain; the N-terminal domains are more similar (14). Nevertheless, even NAC_E100-his_6 is much more insoluble than NAC_K100-his_6. Thus, future studies will continue to focus on NAC_K for in vitro studies.

ACKNOWLEDGMENT

This work was supported by Public Health Service grant GM47156 from the National Institutes of Health to R.A.B.

REFERENCES

1. **Adams, G. M., and R. A. Bender.** Unpublished observations.
2. **Bender, R. A.** 1991. The role of the NAC protein in the nitrogen regulation of *Klebsiella aerogenes*. Mol. Microbiol. 5:2575–2580.
3. **Best, E. A., and R. A. Bender.** 1990. Cloning of the *Klebsiella aerogenes nac* gene which encodes a factor required for nitrogen regulation of the histidine utilization (*hut*) operons in *Salmonella typhimurium*. J. Bacteriol. 172:7043–7048.
4. **Churchward, G., D. Belin, and Y. Nagamine.** 1984. A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. Gene 31:165–171.
5. **Deich, R. A., B. J. Metcalf, C. W. Finn, J. E. Farley, and B. A. Green.** 1988. Cloning of genes encoding a 15,000-dalton peptidoglycan-associated outer membrane lipoprotein and an antigenically related 15,000-dalton protein

from *Haemophilus influenzae*. J. Bacteriol. **170**:489–498.
6. **Feng, J., T. J. Goss, R. A. Bender, and A. J. Ninfa.** 1995. Activation of transcription initiation from the *nac* promoter of *Klebsiella aerogenes*. J. Bacteriol. **177**:5523–5534.
7. **Feng, J., T. J. Goss, R. A. Bender, and A. J. Ninfa.** 1995. Repression of the *nac* promoter of *Klebsiella aerogenes*. J. Bacteriol. **177**:5535–5538.
8. **Goss, T. J., and R. A. Bender.** 1995. The nitrogen assimilation control protein, NAC, is a DNA binding transcription activator in *Klebsiella aerogenes*. J. Bacteriol. **177**:3546–3555.
9. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. Gene **28**:351–359.
10. **Janes, B. K., and R. A. Bender.** 1998. Alanine catabolism in *Klebsiella aerogenes*: molecular characterization of the *dadAB* operon and its regulation by the nitrogen assimilation control protein. J. Bacteriol. **180**:563–570.
11. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
12. **Macaluso, A., E. A. Best, and R. A. Bender.** 1990. Role of the *nac* gene product in the nitrogen regulation of some NTR-regulated operons of *Klebsiella aerogenes*. J. Bacteriol. **172**:7249–7255.
13. **Magasanik, B.** 1996. Regulation of nitrogen utilization, p. 1344–1356. *In* F. C. Neidhardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
14. **Muse, W. B., and R. A. Bender.** 1998. The *nac* (nitrogen assimilation control) gene from *Escherichia coli*. J. Bacteriol. **180**:1166–1173.
15. **Schell, M. A.** 1993. Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. **47**:597–626.
16. **Schwacha, A., and R. A. Bender.** 1993. The *nac* (nitrogen assimilation control) gene from *Klebsiella aerogenes*. J. Bacteriol. **175**:2107–2115.
17. **Schwacha, A., and R. A. Bender.** 1993. The product of the *Klebsiella aerogenes nac* (nitrogen assimilation control) gene is sufficient for activation of the *hut* operons and repression of the *gdh* operon. J. Bacteriol. **175**:2116–2124.
18. **Tyrrell, R., K. H. G. Verschueren, E. J. Dodson, G. N. Murshudov, C. Addy, and A. J. Wilkinson.** 1997. The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. Structure **5**:1017–1032.

Eur. J. Biochem. 214, 17–25 (1993)

Molecular cloning and nucleotide sequence of cDNA encoding rat kidney long-chain L-2-hydroxy acid oxidase

Expression of the catalytically active recombinant protein as a chimaera

Ahmed BELMOUDEN[1], K. H. Diêp LÊ[1], Florence LEDERER[1] and Henri-Jean GARCHON[2]

[1] Centre National de la Recherche Scientifique, Unité de Recherche Associée 1461, Clinique Néphrologique, Hôpital Necker, Paris, France
[2] Institut National de la Santé et de la Recherche Médicale, U25, Clinique Néphrologique, Hôpital Necker, Paris, France

(Received November 30, 1992/January 27, 1993) – EJB 92 1714

Long-chain L-α-hydroxy acid oxidase from rat kidney is a member of the family of FMN-dependent α-hydroxy-acid-oxidizing enzymes. With the knowledge of the recently determined amino acid sequence, the cDNA encoding the enzyme has now been cloned using the polymerase chain reaction. The 1648-bp cDNA contains an open reading frame coding for the 352 residues of the previously determined sequence, preceded by a methionine codon. In addition, several clones were found to present a nine-base insertion, predicting the existence of an isoform with a tripeptide VRK inserted between residues 188 and 189 of the mature protein. The presence of about 10% of this isoform in the oxidase purified from rat kidney was indeed identified by amino acid sequencing. A recombinant active enzyme was obtained as a protein fused to glutathione *S*-transferase using the bacterial expression plasmid pGEX-3X. Physico-chemical characterization indicated, for the fused enzyme, properties similar to those of the rat kidney protein. When the chimaera was submitted to factor Xa, proteolysis at the engineered cleavage point was poor. Separation of hydroxy acid oxidase from glutathione *S*-transferase could not be achieved with trypsin either. With both proteases, the initial cleavage point appeared to be in a peptide loop internal to the hydroxy acid oxidase sequence, close to or in the tripeptide insertion locus and not at the engineered factor-Xa-cleavage point. Comparative tryptic proteolysis of the rat kidney enzyme yielded a form cleaved in the same loop.

Long-chain α-hydroxy acid oxidase is an FMN-dependent dehydrogenase-oxidase, first described by Blanchard et al. (1946). It preferentially oxidizes long-chain aliphatic and aromatic L-α-hydroxy acids in contrast to isozyme A, glycolate oxidase, which preferentially oxidizes the shortest α-hydroxy acid. The rat kidney enzyme was first isolated as an enzyme with weak L-amino acid oxidase activity (Blanchard et al., 1945). More recently, Brush and Hamilton (1981) suggested thiolglyoxylate adducts were its true physiological substrates, even though these compounds are not processed at a higher rate than hydroxybutyrate. This peroxisomal enzyme is found together with isozyme A in hog, beef and sheep kidney (Robinson et al., 1962; McGroarty et al., 1974; Zaar and Fahimi, 1991), whereas it represents the only α-hydroxy-acid-oxidizing activity in rat kidney (Blanchard et al., 1946; Nakano and Danowski, 1966; McGroarty et al., 1974; Angermüller et al., 1986; Zaar and Fahimi, 1991).

The recently determined amino acid sequence of rat kidney long-chain hydroxy acid oxidase (HAO) (Lê and Lederer, 1991) established homology between this protein and other FMN-dependent α-hydroxy-acid-oxidizing enzymes, including glycolate oxidase from spinach (Volokita and Somerville, 1987; Cederlund et al., 1988), lactate oxidase from *Mycobacterium smegmatis* (Giegel et al., 1990), mandelate dehydrogenase from *Pseudomonas putida* (Tsou et al., 1990) and flavocytochrome b_2 or L-lactate dehydrogenase from yeasts (Lederer et al., 1985; Guiard, 1985; Black et al., 1989; Risler et al., 1989). The reaction catalyzed by all these enzymes occurs in two stages. In the reductive half-reaction, the substrate is oxidized and FMN reduced; in the oxidative half-reaction, the flavin is reoxidized. Mechanistic studies, combined with sequence comparisons and the knowledge of the three-dimensional structures of glycolate oxidase and flavocytochrome b_2 (Lindqvist, 1989; Xia and Mathews, 1990; Lindqvist et al., 1991), have led to the idea that the flavin

Correspondence to F. Lederer, Unité de Recherche Associée 1461, Clinique Néphrologique, Hôpital Necker, F-75743 Paris Cedex 15, France
Fax: +33 1 43062388.

Abbreviations. PCR, polymerase chain reaction; HAO, long-chain L-2-hydroxy acid oxidase; ss, single stranded; IPTG, isopropyl β-D-thiogalactopyranoside; Cl_2Ind, 2,6-dichlorophenol indophenol; GST, glutathione S-transferase; ANC, anchor adaptor.

Enzymes. L-amino acid oxidase [(*S*)-amino acid:oxygen 2-oxidoreductase] (EC 1.4.3.2); factor Xa (EC 3.4.21.6); glutathione *S*-transferase (EC 2.5.1.18); glycolate oxidase [(*S*)-hydroxy acid:oxygen 2-oxidoreductase, isozyme A] (EC 1.1.3.15); L-hydroxy acid oxidase [(*S*)-hydroxy acid:2-oxygen oxidoreductase; isozyme B] (EC 1.1.3.15); (*S*)-lactate:cytochrome *c* 2-oxidoreductase or flavocytochrome b_2 (EC 1.1.2.3); L-lactate oxidase [(*S*)-lactate:oxygen 2-oxidoreductase (decarboxylating)] (EC 1.13.12.4); trypsin (EC 3.4.21.4).

Note. The novel nucleotide sequence data published here have been submitted to the EMBL, GenBank and DDBJ sequence data banks and are available under the accession number X67156.

reductive half-reaction has a very similar, if not identical, mechanism for all these enzymes (Ghisla, 1982; Urban and Lederer, 1985; Lederer and Mathews, 1987; Urban et al., 1988; Lindqvist and Brändén, 1989; Giegel et al., 1990; Lederer, 1992). In contrast, the oxidative half-reaction is not identical; the oxidases utilize oxygen with production of hydrogen peroxide, whereas the flavocytochrome b_2 flavin has a built-in monoelectronic acceptor, heme b_2. The physiological acceptor for mandelate dehydrogenase is not known. Clues as to the structural differences which lead to the differentiation between the dehydrogenase/oxidase and the dehydrogenase/electron transferase function were suggested by the comparison between the three-dimensional structures of flavocytochrome b_2 and glycolate oxidase (Lindqvist et al., 1991).

Nowadays, one efficient way of furthering the understanding of structure/function relationships is the mechanistic comparison between wild-type enzymes and site-directed mutants. Such a study is underway for flavocytochrome b_2 from *Saccharomyces cerevisiae* (Reid et al., 1988; Dubois et al., 1990; Miles et al., 1992). In view of a number of kinetic similarities between this enzyme and HAO (Cromartie and Walsh, 1975b; Pompon et al., 1980; Urban and Lederer, 1984; Urban et al., 1988), we decided to use this same approach with HAO in order to pinpoint the requirements for flavin reactivity with oxygen. As a first step in this direction, we report here the cloning of the cDNA encoding HAO from rat kidney and its expression in *Escherichia coli* as a catalytically active enzyme fused with glutathione *S*-transferase.

MATERIALS AND METHODS

Materials

The plasmid pKS M13+/−, *E. coli* strains XL1-blue and HB 101 were from Stratagene. Expression plasmids pGEX-3X and pKK233-3 were from Pharmacia. All restriction endonucleases, modifying enzymes and factor Xa were from New England Biolabs, while Mu-MLV reverse transcriptase, terminal deoxynucleotidyl transferase and *Taq* DNA polymerase were from Bethesda Research Laboratories. Tosylphenylalanine-chloromethane-treated trypsin was from Merck. DNA sequencing was performed on double-stranded DNA using modified T7 DNA polymerase (sequenase) from U. S. Biochemicals. [α-^{32}P] dCTP and [α-^{35}S] dATP were purchased from Amersham Corporation. All other chemicals were of reagent grade. Oligodeoxynucleotides were synthesized on a model 391 Applied Biosystems DNA synthesizer.

Isolation of mRNA and Northern-blot analysis

Total RNA was extracted from fresh kidney tissue from one Wistar rat with the acid/phenol method (Chomczynski and Sacchi, 1987). Poly(A)-rich RNA was isolated from total RNA by oligo(dT)-cellulose chromatography (Aviv and Leder, 1972). Total and poly(A)-rich RNA were subjected to electrophoresis on a 0.8% agarose/formaldehyde gel (Lehrach et al., 1977), transferred to a nitrocellulose filter, hybridized with the 362-bp polymerase chain reaction (PCR) fragment and radiolabelled by random priming (Feinberg and Vogelstein, 1983). Hybridization was at 65°C in Church's solution (Church and Gilbert, 1984). Washing was at a final stringency of 0.1X solution A at 60°C (20X solution A is 3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA, pH 7.4).

Internal-cDNA amplification

Single-stranded (ss) cDNA was synthesized from poly-(A)-rich kidney RNA (5 µg) using Mu-MLV reverse transcriptase and oligo-$(dT)_{15}$ as a primer. Amplification, using the 'mixed oligonucleotide primed amplification of cDNA' procedure (Lee et al., 1988; Girgis et al., 1988), was carried out in 50 µl with 100 ng ss cDNA and 200 pmol each primer mixture (X1 and X2). After 30 cycles (94°C, 1 min; 37°C, 2 min; 72°C, 1.5 min), the reaction mixture was fractionated by electrophoresis (1% low-melting-point agarose in 40 mM Tris/acetate, 1 mM EDTA, pH 7.2) and aliquots from five gel slices, corresponding to the 200−400-bp region, were further amplified in the same conditions for 30 cycles.

5′-end amplification

dG-tailing of ss cDNA was performed according to Dent and Wu (1984). Amplification was carried out in 50 µl with 8 ng dG-tailed ss cDNA, using gene-specific X3 (40 pmol) and oligo-$(dC)_{15}$ (200 pmol). After 30 cycles (94°C, 1.5 min; 45°C, 1.5 min; 72°C, 1.5 min), the reaction product was isolated by electrophoresis (1% low-melting-point agarose) and an aliquot was reamplified with gene-specific X4 and oligo-$(dC)_{15}$ (5 pmol each) for 30 cycles as described above.

3′-end amplification

5 µg poly(A)-rich RNA was reverse transcribed in 50 µl as described above, except for substitution of 45 pmol $(dT)_{17}$-ANC to oligo-$(dT)_{15}$ primer. The reaction mixture was diluted to 1 ml with solution B (10 mM Tris/HCl, 1 mM EDTA, pH 7.5) and stored at 4°C. The cDNA pool (5 µl) was amplified in 50 µl with gene-specific X5 and an anchor adaptor (ANC) as primers (25 pmol each), using 55°C as annealing temperature and 1.5 min for the extension step at 72°C (Frohman and al., 1988). Two additional rounds of amplification were performed with the nested X6 and X7 primers for the 5′-end and ANC for the 3′-end.

Cloning and sequencing of cDNA

PCR products were extracted with phenol/chloroform and ethanol precipitated. The Klenow fragment and T4 polynucleotide kinase were used to fill-in ends and phosphorylate 5′-termini respectively (Sambrook et al., 1989).

After electrophoresis (1% low-melting-point agarose), the region of the gel containing specific products was isolated and the DNA was cloned in the *Eco*RV or *Sma*I site of the Bluescript vector. Double-stranded DNA from recombinant plasmids was prepared using standard procedures (Sambrook et al., 1989) and sequenced with modified T7 DNA polymerase (sequenase), following the dideoxy-chain-termination procedure (Sanger et al., 1977) according to the recommendations of the supplier.

Attempts at direct expression of HAO cDNA

The entire coding region of HAO was obtained by PCR amplification, using suitable primers, of the ss cDNA obtained from poly(A)-rich RNA. It was then cloned into the *Nco*I site of the prokaryotic expression vector pKK233-3. The plasmid was introduced into *E. coli* strain JM 105. After induction with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) under standard conditions, no band was observed at

the expected position on the Coomassie-blue-stained protein pattern of a SDS/gel electrophoresis. Changing induction conditions, or growth temperature, and introducing the plasmid into other strains (HB 101, XL1-blue, AR 120, N5151) were unsuccessful.

Construction of the expression plasmid carrying rat hydroxy acid oxidase cDNA fused to glutathione *S*-transferase cDNA

Procaryotic expression vector pGEX-3X (Smith and Johnson, 1988) was linearized with *Sma*I and treated with calf intestinal alkaline phosphatase. This vector was then ligated to the entire coding region of HAO cDNA obtained after PCR amplification. The recombinant plasmid (pGHAOX) was introduced into *E. coli* strain XL1-blue.

Expression of hydroxy acid oxidase

E. coli transformed with pGHAOX was cultured overnight in Luria-Bertani medium containing 50 μg/ml ampicillin and 12.5 μg/ml tetracycline at 37°C. The culture was then diluted 1:50 into 4 l of the same medium and incubated at 37°C with good aeration. When the absorbance, measured at 600 nm, reached a value of 0.4, IPTG was added to a final concentration of 1 mM and incubation was continued for 3 h. Aliquots of cells (1 ml) were harvested before and after induction, pelleted, resuspended in 100 μl SDS-loading buffer, then subjected to 0.1% SDS/10% PAGE (Laemmli, 1970). Protein bands were visualized by Coomassie-blue staining.

Purification of hydroxy acid oxidase

Rat kidney enzyme was purified as described by Urban et al. (1988). For obtaining the recombinant enzyme, frozen cells from a 4-l culture, grown as described above, were thawed and lysed by sonication in 40 ml NaCl/P_i-Triton (150 mM NaCl, 20 mM sodium phosphate, 1% Triton X-100, pH 7.3). After centrifugation, the supernatant was loaded on a glutathione–Sepharose-4B column (4 ml total volume) under conditions recommended by the manufacturer. After elution with 5 mM glutathione in 50 mM Tris/HCl, pH 8, the fusion protein (300 nmol) was precipitated by addition of ammonium sulphate to 70% saturation; the precipitate was dissolved in a minimal volume of 50 mM Tris/HCl, 10 μM FMN, pH 8.0, dialyzed against the same buffer and stored at −80°C.

Molecular-sieve chromatography

An AcA34 column (1.6 cm × 80 cm) was equilibrated in 100 mM sodium/potassium phosphate buffer, 1 mM EDTA, pH 7, at a flow rate of 15 ml/h. A mixture of 1.2 mg flavocytochrome b_2 [M_r 230000 (Lederer et al. 1985)], 2.7 mg glutathione-*S*-transferase–HAO chimaera (GST, glutathione *S*-transferase) and 0.9 mg rat kidney hydroxy acid oxidase [M_r 156000 (Lê and Lederer, 1991)] was chromatographed on the column at the same flow rate. The effluent was monitored at 280 nm. Protein-containing fractions were assayed for mandelate/Cl_2-Ind oxidoreductase activity; (HAO activity; Cl_2Ind, 2,6-dichlorophenol indophenol) and lactate-ferricyanide oxidoreductase activity (flavocytochrome b_2). HAO can oxidise lactate but its specific activity with this substrate is more than 250-fold lower than that of flavocytochrome b_2; the latter cannot oxidize mandelate (Lederer, 1991).

Enzymic assays

The standard activity assay solution consisted of 25 mM D,L-2-hydroxybutyrate, 70 μM Cl_2-Ind (0.002%) in 0.1 M Tris/HCl, pH 8.4, and 1 mM EDTA. Dye reduction was monitored with a Uvikon 930 spectrophotometer at 30°C. An absorption coefficient of 22000 M^{-1} cm^{-1} was used for Cl_2Ind at 600 nm. The enzyme activity unit is expressed in nmol dye reduced · s^{-1} · mol enzyme^{-1}. Protein concentrations were determined by the Bio-Rad protein microassay procedure using bovine serum albumin as a standard. The enzyme concentration was expressed relative to its FMN content (ε_{450} = 11700 M^{-1} cm^{-1}).

Proteolysis with trypsin and factor Xa

Proteolysis with factor Xa was performed in the buffer described by Nagai and Thøgersen (1984), at a protease/substrate ratio of 1:100. Proteolysis with trypsin was performed in 50 mM Tris, pH 8, at a protease/substrate ratio of 1:500 at 0°C. At each given time, aliquots were removed and quenched with 1 vol. 60 μM tosyllysyl chloromethane in 100 mM imidazole/HCl, pH 7.5.

Peptide sequencing

SDS/PAGE was carried out in Tris/borate (45 mM Tris, 44 mM boric acid, 2 mM EDTA, 0.1% SDS, pH 7). Electrotransfer was carried out in the buffer in the absence of SDS, with the semidry apparatus from Millipore (Miniblot-SDE) at 0.2 mA/cm^2 for 1 h. Automated Edman degradation was carried out on the excised spots with a gas/liquid-phase sequencer model 477A from Applied Biosystems equipped with on-line phenylthiohydantoin amino acid analysis.

RESULTS

Isolation of the cDNA of rat kidney hydroxy acid oxidase

Knowledge of the amino acid sequence of the enzyme (Lê and Lederer, 1991) was used for synthesizing two sets of degenerate oligonucleotides, X1 and X2 (Table S1), as primers for amplifying a 362-bp internal fragment of ss cDNA from rat kidney, using PCR (Figs S1 and S2). After cloning and sequencing, the amplified DNA showed an open reading frame corresponding to the expected amino acid sequence. Non-degenerate oligonucleotides, corresponding to various parts of this internal sequence, were then chosen to obtain the full-length cDNA sequence following the anchored PCR protocol (Loh et al., 1989; Table S1 and Figs S1 and S2). For 5′-end amplification, X3 and $(dC)_{15}$ were used after dG-tailing of the ss cDNA. The product was further amplified using $(dC)_{15}$ and the nested primer X4 in order to increase specificity. The resulting 370-bp fragment was cloned and sequenced (three independent clones). For the 3′-end amplification, poly(A)-rich mRNA was first reverse transcribed using as a primer $(dT)_{17}$-ANC (Table S1), according to Frohman et al. (1988) (the ANC adaptor contains the *Xho*I, *Sal*I and *Cla*I restriction sites). The ss cDNA was then submitted to three rounds of amplification using the nested oligonucleotides X5, X6 and X7 successively for priming the 5′-side and ANC for the 3′-side. Two fragments of about 1.1

Table 1. Comparison of kinetic parameters of the glutathione-*S*-transferase–hydroxy-acid-oxidase chimaera and of rat kidney hydroxy acid oxidase. The kinetic determinations were carried out in 0.1 M imidazole/HCl, pH 7.5 at 30°C. n.d., not determined.

Enzyme	D,L-2-Hydroxybutyrate			L-Mandelate		
	K_m	k_{cat}	k_{cat}/K_m	K_m	k_{cat}	k_{cat}/K_m
	mM	s^{-1}	$mM^{-1}\ s^{-1}$	mM	s^{-1}	$mM^{-1}\ s^{-1}$
GST-HAO	2.25 ± 0.1	1.66 ± 0.03	0.74	0.42 ± 0.01	1.89 ± 0.02	4.5
HAO (this work)	0.69 ± 0.02	1.66 ± 0.02	2.40	n.d.	n.d.	n.d.
HAO (Urban et al., 1988)	1.0 ± 0.1	1.02 ± 0.02	1.0	0.40 ± 0.02	1.52 ± 0.08	3.8

kb were obtained, subcloned and sequenced, using the strategy shown in Fig. S3.

Fig. S2 shows the nucleotide sequence deduced for the cDNA of the HAO gene by merging the information obtained from sequencing the three PCR fragments; the 5′-end, the central part and the lower-molecular-mass 3′-end. The sequence contains a 5′-untranslated region of 202 bp, an open reading frame of 1059 bp and a 3′-untranslated region of 387 bp [excluding the poly(A)-rich tail]. The coding region corresponds to the expected HAO amino acid sequence, with an additional methionine before Pro1 of the mature protein. As no other ATG codon is found further upstream in the correct reading frame, this must be the initiator codon, even though, of the consensus sequence for the initiation of protein synthesis in eukaryotes, only G at position −3 and C at position −5 are present (Kozak, 1984, 1986). The methionine must be removed by proteolysis during biosynthesis. No N-terminal signal sequence is expected for HAO, which possesses the C-terminal peroxisomal targeting sequence SRL (Gould et al., 1990).

The higher-molecular-mass PCR fragment of about 1.1 kb corresponding to the 3′-end had, after position 568 for the three clones that were examined, a nine-base insertion corresponding to TGAGAAAGG. A nine-base 3′-terminal extension (AAAAATGAA) was also present. The nine-base insertion corresponds to a 3 amino acid insertion (VRK) between E188 and E189 of the mature protein. The presence of these residues had not been detected in the amino acid sequence work (Lê and Lederer, 1991).

Tissue distribution of hydroxy acid oxidase mRNA

Northern-blot analysis of rat kidney poly(A)-rich RNA, using, as a probe for hybridization under stringent conditions, the 362-bp internal PCR fragment obtained with X1 and X2, revealed the presence of two strong bands of ≈1.8 kb and 2.0 kb (Fig. 1A). These bands were also easily visualized in total kidney RNA. No signal was observed in total RNA samples from other rat tissues (lung, heart, brain, muscle, liver and testis; Fig. 1B). This confirmed the tissue specificity of the kidney enzyme (Tolbert, 1981). The length of the cloned cDNA (1648 bp) suggested it corresponded to the lower-molecular-mass mRNA.

Expression of hydroxy acid oxidase in *E. coli* as a fusion protein with glutathione *S*-transferase

The full-length coding region, starting at the Pro1 codon, was amplified by PCR from total kidney ss cDNA using primers corresponding to the 5′-end and 3′-end of the coding sequence (20 nucleotides). Its sequence was checked and

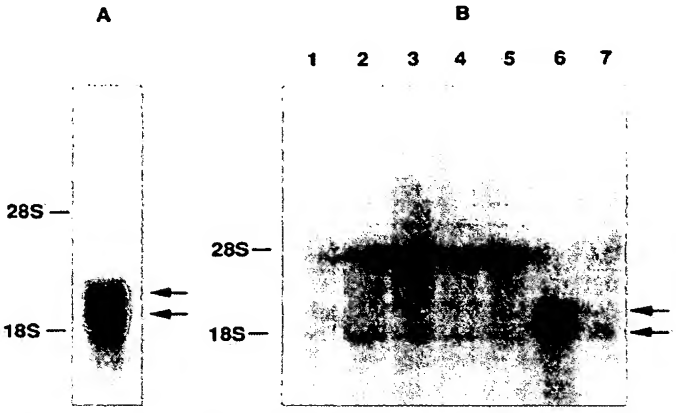


Fig. 1. Size and tissue distribution of hydroxy acid oxidase mRNA. 5 μg poly(A)-rich RNA from rat kidney (A) and 10 μg total RNA from different rat tissues (B) were analysed by Northern-blot analysis using, as a probe, the 362-bp fragment obtained with X1 and X2 as primers. In (B), RNA was from lung (1), heart (2), brain (3), muscle (4), liver (5), kidney (6) and testis (7). The position of the 18S and 28S ribosomal RNA is indicated. The arrows highlight the two bands observed for hydroxy acid oxidase mRNA.

found to correspond to that shown in Fig. S2, i.e. it did not present the VRK insertion. Initial attempts to express HAO as an unfused protein using the pKK233−3 bacterial expression vector were unsuccessful, despite a correct sequence and the use of five different *E. coli* strains. As an alternative, the HAO coding region was subcloned into the prokaryotic expression vector pGEX-3X. The resulting construct is expected to yield a fusion protein with GST from *Schistozoma japonicum*. Moreover, the chimaeric protein has been designed to contain a specific proteolytic cleavage site for bovine factor Xa, between the N-terminal GST sequence and the protein to be expressed (IEGR/GIP etc). Insertion of the HAO coding region at the *Sma*I site, combined with factor-Xa cleavage of the fusion protein, should thus yield the oxidase with an N-terminal extension of three residues (GIP) before Pro 1 of the natural protein. The recombinant plasmid, pGHAOX, was introduced into *E. coli* strain XL1-blue. A number of clones were examined for protein expression after induction with IPTG. Their total protein content was analyzed by SDS/PAGE and Coomassie-blue staining after cell lysis. Fig. 2A shows a typical example of a clone overexpressing a 65-kDa protein found only after IPTG treatment of the bacterial culture. The protein was absent from the cells transformed with the original pGEX-3X, which conversely expressed a 26-kDa protein corresponding to the expected

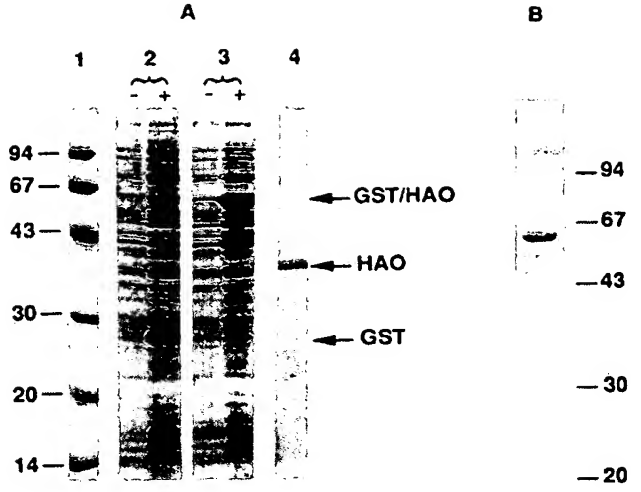


Fig. 2. Expression of rat kidney hydroxy acid oxidase in *E. coli*. (A) Total cell proteins on 12.5% SDS/PAGE. 1-ml cultures were harvested before (−) and after (+) induction with 1 mM IPTG for 3 h; the pellet was dissolved in loading buffer, and one half was electrophoresed after boiling. Lane 1, molecular-mass markers; lane 2, XL1-blue transfected with pGEX-3X; lane 3, XL1-blue transformed with pGHAOX; lane 4, hydroxy acid oxidase purified from rat kidney. (B) Purified fusion protein after chromatography on glutathione−Sepharose 4B.

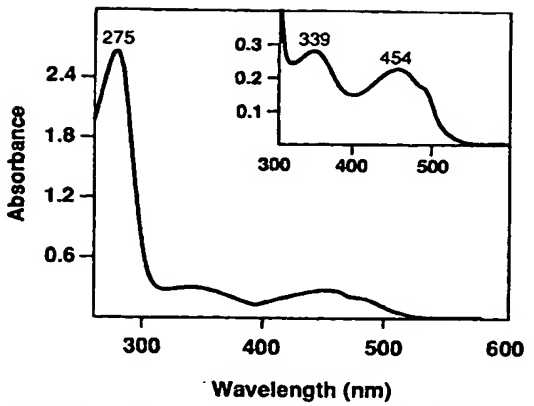


Fig. 3. Absorption spectrum of the purified glutathione *S*-transferase/hydroxy acid oxidase chimaera (21 μM enzyme in 50 mM Tris/HCl, pH 8.0). The insert shows the visible spectrum at a higher sensitivity.

GST. The recombinant fusion protein expressed by cells transfected with pGHAOX had the size expected for the GST-HAO fusion protein, and it represented about 5% of the total protein content.

Enzymic and physicochemical characterization of the GST/HAO chimaera

The fusion protein was purified by affinity chromatography on glutathione−Sepharose-4B. Cells from an IPTG-induced culture were sonicated in the presence of 1% Triton X-100 and the crude supernatant was adsorbed on the affinity column. After all the Triton X-100 and most of the protein had been washed out from the column, elution was carried out with glutathione in the absence of detergent. This indicated that the fusion protein was soluble. Coomassie-blue staining of an SDS/polyacrylamide gel showed a single protein band with the expected M_r of 65000 (Fig. 2B). Upon chromatography on an AcA34 molecular-sieve column (see Materials and Methods), the HAO activity peak was eluted one fraction volume ahead of the lactate-oxidizing activity of flavocytochrome b_2. Since the latter is a tetramer of M_r 4×57500, this indicated a M_r for the chimaera compatible with a tetrameric structure (M_r 4×65000). The enzyme solution was yellow, which indicated the presence of oxidized flavin. Its absorption spectrum showed the two characteristic flavin peaks, with maxima at 454 nm and 339 nm (Fig. 3). Relative to the spectrum of the enzyme purified from rat kidney, the latter peak presented a slight blue shift, probably due to interference by the tailing of the higher ultraviolet peak at 275 nm (arising from the GST chromophores).

The enzymic parameters of the fusion protein were determined for two substrates: D,L-2-hydroxybutyrate and L-mandelate. The results are compared in Table 1 to those obtained with the enzyme purified from rat kidney. While there is excellent agreement with published values relative to mandelate (Urban et al., 1988), small differences were observed with α-hydroxybutyrate. The values obtained in this work with the rat kidney enzyme are no more different from those of Urban et al. (1988) than what can be expected from experiments using different enzyme batches. However, there is a small, but real difference, in the K_m for hydroxybutyrate between the fusion protein and the wild-type enzyme.

Proteolytic cleavage of the fusion protein with factor Xa and with trypsin

When factor Xa was used, from two different commercial sources, no matter what the molar ratio of protease/substrate was (1:100 and 1:10), little proteolysis took place (Fig. 4A). After a 24-h digestion, the SDS/polyacrylamide gel showed, besides a majority of intact fusion protein, two faint bands with apparent M_r 43000 and 18000, similar to those obtained with trypsin. It thus appeared that the engineered factor-Xa cleavage site (IEGR/GIP) was not accessible to the protease in the fusion protein.

At 0°C with a trypsin/ substrate ratio of 1:500, the fusion protein was rapidly digested and three main bands were obtained on the SDS/polyacrylamide gel at 43, 26 and 18 kDa (the last one called β; Fig. 4B). Amino acid sequencing of the fragments after electroblotting showed fragment β to correspond to the C-terminal part of HAO, beginning at Ala185. The other two fragments presented the N-terminal sequence of engineered GST (MSPILGY, etc.). This suggested that two areas had been cleaved by trypsin in the fusion protein; the engineered IEGR/GIP sequence between the two domains and a region in the HAO sequence also cleaved by rat kidney proteases during kidney handling or during enzyme purification (Lê and Lederer, 1991). The relative abundance of the higher-molecular-mass band suggested that the HAO-sensitive region was cleaved first, producing the 43-kDa fragment (GST plus the 184-residue N-terminal HAO fragment α) and an 18-kDa fragment β (HAO residues 185−352). The 43-kDa fragment could then be cleaved at the factor-Xa cleavage point, yielding protease-resistant GST and a protease-sensitive fragment α. Alternatively, after ini-

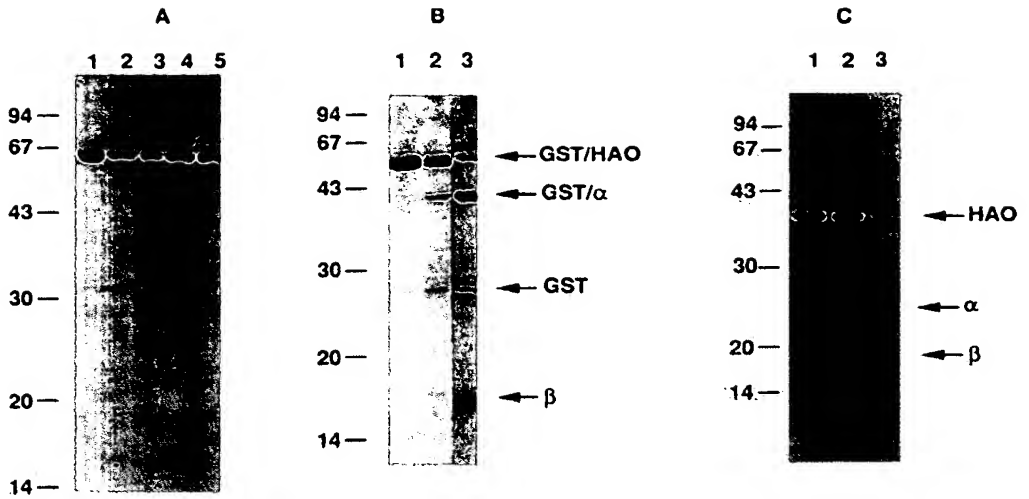


Fig. 4. Proteolysis of the glutathione *S*-transferase/hydroxy acid oxidase chimaera (A and B) and of rat kidney hydroxy acid oxidase (C). (A) Proteolysis of the fusion protein with factor Xa (protease/substrate molar ratio 1:15; substrate, 25−30 μM; 50 mM Tris/HCl, 100 mM NaCl, 4 mM $CaCl_2$, pH 8.0, 25°C). Lane 1, zero time; lanes 2 and 3, 30 min and 120 min, factor Xa from Biolabs; lanes 4 and 5, factor Xa from Sigma Chemical Co (contains bovine serum albumin which gives rise to the second Coomassie-blue-stained band). (B) Proteolysis of the fusion protein with trypsin (protease/substrate molar ratio, 1: 500; substrate, 25−30 μM; 50 mM Tris/HCl, pH 8, 0°C). Lane 1, zero time; lane 2, 5 min; lane 3, 2 h. (C) Proteolysis of HAO purified from rat kidney. Same conditions as for (B).

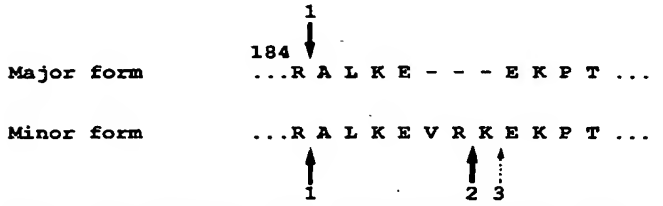


Fig. 5. Tryptic cleavage in the protease-sensitive region of HAO purified from rat kidney. Arrows indicate the location of the cleavage points as deduced from amino acid sequencing results of fragment *β*.

tial HAO loop cleavage, the GST *α* fragment could possibly be digested piecemeal yielding GST as a stable end product.

For verification of the conformational integrity of the HAO domain in the fusion protein, the purified rat kidney enzyme was submitted to trypsin digestion under conditions identical to those used for the chimaera. Both fragments *α* (a doublet at 22 kDa) and *β* (18 kDa) were formed (Fig. 4C), confirming that the natural protein contains a protease-sensitive loop and suggesting that fragment *α* is more sensitive to further digestion by trypsin when it is fused to GST than when it is not. The two bands corresponding to fragment *α* had the same N-terminal sequence (PLV, etc.) after electroblotting, showing their difference to lie at their C-termini. Amino acid sequencing of fragment *β* yielded two sequences, and three in a case of prolonged digestion. The major sequence (ALKEEKP) showed the tryptic cleavage point to lie between R184 and A185, as with the fusion protein (Fig. 5). The minor sequence, which represented about 10−15% of the total amount sequenced, was KEKPTQSV. This can only have originated from a protein possessing the VRK insertion, which would be cleaved between R and K (Fig. 5). In the case of the prolonged digestion, the major sequence was identical. The minor sequence was EKPTQ, etc.; it could have arisen from a cleavage after the lysine of the VRK insertion (Fig. 5), but the idea of a less probable but not impossible tryptic cleavage between E188 and E189 of the main sequence cannot be totally ruled out (Haumont et al., 1989 and references therein). It can be added that enzymic activity monitoring during proteolysis showed HAO activity to decrease by about 15% over the first 100 min, after which it was as stable as that of the control. This suggested that proteolytic cleavage had only a slight effect on the HAO enzymic activity.

DISCUSSION

Using the information given by the amino acid sequence of *α*-hydroxy acid oxidase, we used PCR for amplifying, cloning and sequencing three overlapping fragments from rat kidney cDNA, namely a central coding segment, a 5′ fragment and a longer 3′ end. Several independent clones were sequenced each time in order to detect possible copying errors of *Taq* DNA polymerase. Most of the clones gave the sequence shown in Fig. S2, yielding an open reading frame identical to the amino acid sequence previously determined. However, three out of eight clones of the 3′-fragment carried the same nine-base insertion in the open reading frame, which would correspond to the insertion of amino acids VRK between E188 and E189 of the determined amino acid sequence. This cannot be ascribed to a copying error. The three-residue insertion occurs in a region which is sensitive to proteases, as shown in this and previous work (Lê and Lederer, 1991), and corresponds to disordered loops in the three-dimensional structures of the homologous enzymes flavocytochrome b_2 (Xia and Mathews, 1990) and glycolate oxidase (Lindqvist, 1989). Thus, this region could probably accommodate a few more residues without alteration of the overall structure. No other sequence difference was detected in the cDNA, apart from an extension of the poly(A) tail. These results suggested the possible existence of *α*-hydroxy acid oxidase isozymes. While no evidence for amino acid

sequence heterogeneity was obtained before (Lê and Lederer, 1991), during the present work the tryptic cleavage in the loop region of kidney HAO followed by automatic Edman degradation provided evidence for the existence of a second sequence compatible with the insertion of the tripeptide VRK. The corresponding isoform was calculated to represent about 10% of the total protein submitted to degradation. In the past, McGroarty et al. (1974) and Cromartie and Walsh (1975a) noted the presence of one major and 1−3 additional faint bands staining for HAO activity on non-denaturing gels. The molecular nature of these putative isozymes was not elucidated. In contrast, Duley and Holmes (1974) detected only one activity band on starch gels. Thus, all the evidence obtained to date suggests that a second form of HAO, if present, must be minor.

Alternative splicing of the primary transcript from a single gene could possibly explain the generation of these two HAO isoforms, given the sequence identity for the untranslated regions of their mature mRNA. The existence of the minor isoform at the protein level is intriguing from an evolutionary point of view. One of the major differences between the various members of the FMN-dependent α-hydroxy-acid-oxidizing enzymes lies in the length of the polypeptide chain which loops out of the $\beta_8\alpha_8$ barrel between strand β_4 and helix α_4 (Lê and Lederer, 1991). It is in this segment that the structurally disordered region is found, as well as the protease-sensitive bonds and the tripeptide insertion. It is so far not known whether this region which loops out of the β barrel plays an important role in the structure or function of the enzyme. It was, however, found that proteolytic cleavage of a single peptide bond in flavocytochrome b_2 led to alterations in a number of kinetic parameters (k_{cat}, K_m, K_i for inhibitors), suggesting some kind of interaction, direct or indirect, between the active site and the disordered region, either before or after cleavage (Ghrir and Lederer, 1981). Moreover, the folding of the extension between β_4 and α_4, with respect to the barrel, is very different in glycolate oxidase (Lindqvist, 1989) and in flavocytochrome b_2 (Xia and Mathews, 1990), probably due to the presence of the heme-binding domain. Thus, this region of the peptide chain may be a hot spot for evolutionary changes in the protein family, and it is tempting to speculate that alternative splicing might be a molecular mechanism favouring such changes (Craik et al., 1983). The data presented in this paper do not exclude, however, the existence of closely related genes as an origin of the isoforms. An allelic form of the gene is unlikely because the mRNA used for amplification was extracted from a single homozygous Wistar rat. It is to be noted that a single cDNA sequence was identified for the other known members of the family of hydroxy-acid-oxidizing enzymes.

We used the radiolabeled amplified central fragment for characterising HAO mRNA by Northern blotting. Two messengers of equal abundance and differing size (about 1800 bp and 2000 bp) were detected in rat kidney poly(A)-rich RNA. They were also detected in total kidney RNA; their expression was tissue specific, at least at the sensitivity level of the Northern-blot technique. This tissue specificity is in keeping with previous observations (McGroarty et al., 1974; Zaar and Fahimi, 1991). The origin of the difference between the two messengers remains to be elucidated, since their mass difference cannot be explained by the sequence differences detected in PCR products.

Hydroxy acid oxidase was expressed as a fusion protein with glutathione *S*-transferase using the vector pGEX-3X. After induction with IPTG, up to 5% of the proteins were constituted by the chimaera, expressed as soluble protein. The fusion protein was readily purified by affinity chromatography on immobilised glutathione. That the folding of the chimaeric HAO should be similar if not identical to the native HAO is shown by its flavin spectrum, its enzymic activity and its similar susceptibility to limited proteolysis by trypsin between positions 184 and 185. The enzymic parameters determined for the substrate L-mandelate are identical to those obtained with the enzyme purified from rat kidney. The small difference found for the K_m of D,L-α-hydroxybutyrate is somewhat surprising, as one would expect a structural perturbation due to the fusion to have effects on both substrates. Morcover, it can hardly arise from the presence of the second HAO form in the natural enzyme, since this form is a minor constituent of the purified natural enzyme. Whatever the case, even though we could not obtain intact recombinant HAO due to the resistance of the chimaera to proteolysis, our results open the way to the study of structure/function relationships with mutants whose properties should be compared to those of the wild-type fusion protein.

Mammalian GST enzymes are dimeric (Mannervick and Danielson, 1988), but the enzyme from *Schistosoma japonicum* is believed to be a monomer (Tiu et al., 1988; Farmer et al., 1992). Hydroxy acid oxidase itself is a tetramer (Cromartie and Walsh, 1975b; Duley and Holmes, 1976). All the well characterized homologous enzymes were found to be either tetramers or octamers (Monteilhet and Risler, 1970; Phillips et al., 1976; Sullivan et al., 1977). Furthermore, a monomeric form of *Hansenula anomala* flavocytochrome b_2 obtained at low ionic strength was found inactive, even though it still retained both the flavin and the heme (Baudras, 1972); activity was recovered together with the tetrameric state upon raising the ionic strength. No evidence exists for an active monomeric state for any of the other members of the family. Furthermore, the crystal structures of glycolate oxidase and flavocytochrome b_2 show intersubunit contacts to be extensive (Lindqvist, 1989; Xia and Mathews, 1990; Lindqvist et al., 1991). The GST-HAO chimaera was thus expected, as found, to have a GST domain flanking each subunit of the tetramer. The three-dimensional structures of the homologous enzymes flavocytochrome b_2 and glycolate oxidase (Lindqvist et al., 1991) show an identical fold starting at a position equivalent to mature HAO Leu2 (a turn followed by a helix). A few residues in this area, at the surface of the subunit, are located close to fourfold contacts in the two structures. It would be surprising to find a very different topology at the N-terminus of HAO, in view of its degree of similarity with the two other enzymes. This local structure, therefore, as well as the overall architecture of the tetramer, appear compatible with the formation of a more complex and active molecular edifice, which retains affinity for glutathione and has both flavin-binding capacity and hydroxy acid oxidizing activity, even though the fusion site is inaccessible to factor Xa. It is interesting to note that, while there is a rapid increase in the number of described cases of recombinant proteins or domains expressed as fusion proteins with GST, the number of active enzymes expressed in this way is still rather limited. Furthermore, to our knowledge, the present case represents the first example of an active multimeric enzyme fused to GST.

A. Belmouden was a recipient of a fellowship from the *Centre National de la Recherche Scientifique* (*bourse de doctorat pour ingénieurs*). The authors are indebted to D. Mazella for competent secreterial help.

REFERENCES

Angermüller, S., Leupold, C., Zaar, K. & Fahimi, H. D. (1986) *Histochemistry 85*, 411–418.

Aviv, H. & Leder, P. (1972) *Proc. Natl Acad. Sci. USA 69*, 1408–1412.

Baudras, A. (1972) in *Dynamic aspects of conformation changes in biological macromolecules*, (Sadron, C. ed.) pp. 181–205, D. Reidel Publishing, Dordrecht.

Black, M. T., Gunn, F. J., Chapman, S. K. & Reid, G. A. (1989) *Biochem. J. 263*, 973–976.

Blanchard, M., Green, D. E., Nocito, V. & Ratner, S. (1945) *J. Biol. Chem. 161*, 583–598.

Blanchard, M., Green, D. E., Nocito, V. & Ratner, S. (1946) *J. Biol. Chem. 163*, 137–144.

Brush, E. J. & Hamilton, G. A. (1981) *Biochem. Biophys. Res. Commun. 103*, 1194–1200.

Cederlund, E., Lindqvist, Y., Söderlund, G., Brändén, C. I. & Jörnvall, H. (1988) *Eur. J. Biochem. 173*, 523–530.

Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.

Church, G. M. & Gilbert, W. (1984) *Proc. Natl Acad. Sci. USA 81*, 1991–1995.

Craik, C. S., Rutter, W. J. & Fletterick, R. (1983) *Science 220*, 1125–1129.

Cromartie, T. H. & Walsh, C. T. (1975a) *Biochemistry 14*, 2588–2596.

Cromartie, T. H. & Walsh, C. T. (1975b) *Biochemistry 14*, 3482–3490.

Deng, G. & Wu, R. (1983) *Methods Enzymol. 100*, 96–116.

Dubois, J., Chapman, S. K., Mathews, F. S., Reid, G. A. & Lederer, F. (1990) *Biochemistry 29*, 6393–6400.

Duley, J. & Holmes, R. S. (1974) *Genetics 76*, 93–97.

Duley, J. & Holmes, R. S. (1976) *Eur. J. Biochem. 63*, 163–176.

Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Farmer, K., Catala, F. & Wright, W. E. (1992) *J. Biol. Chem. 267*, 5631–5636.

Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl Acad. Sci. USA 85*, 8998–9002.

Ghrir, R. & Lederer, F. (1981) *Eur. J. Biochem. 120*, 279–287.

Ghisla, S. (1982) in *Flavins and flavoproteins* (Massey, V. & Williams, C. H., eds) pp. 133–142, Elsevier Science Publishers, Amsterdam.

Giegel, D. A., Williams, C. H. & Massey, V. (1990) *J. Biol. Chem. 265*, 6626–6632.

Girgis, S. I., Alevizaki, M., Denny, P., Ferrier, G. J. M. & Legon, S. (1988) *Nucleic Acids Res 16*, 10371.

Gould, S. J., Krisans, S., Keller, G. A. & Subramani, S. (1990) *J. Cell. Biol. 110*, 27–34.

Guiard, B. (1985) *EMBO J. 4*, 3265–3272.

Henikoff, S. (1984) *Gene (Amst.) 28*, 351–359.

Haumont, P. Y., Thomas, M. A., Labeyrie, F. & Lederer, F. (1987) *Eur. J. Biochem. 169*, 539–546.

Kozak, M. (1984) *Nucleic Acids Res. 12*, 857–872.

Kozak, M. (1986) *Cell 44*, 283–292.

Laemmli, U. K. (1970) *Nature 227*, 680–685.

Lê, K. H. D. & Lederer, F. (1991) *J. Biol. Chem. 266*, 20877–20881.

Lederer, F. (1991) in *Chemistry and biology of flavoproteins* (Müller, F., ed.) pp. 153–242, CRC Press Inc.

Lederer, F. (1992) *Protein Science 1*, 540–548.

Lederer, F. & Mathews, F. S. (1987) in *Flavins and flavoproteins* (Edmondson, D. & Mc Cormick, D. B., eds) pp. 133–142, Walter de Gruyter and Co., Berlin.

Lederer, F., Cortial, S., Becam, A. M., Haumont, P. Y. & Perez, L. (1985) *Eur. J. Biochem. 139*, 59–65.

Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M. & Caskey, C. T. (1988) *Science 239*, 1288–1291.

Lehrach, H., Diamond, D., Wozney, J. & Boedtker, H. (1977) *Biochemistry 16*, 4743–4751.

Lindqvist, Y. (1989) *J. Mol. Biol. 209*, 151–166.

Lindqvist, Y. & Brändén, C.-I. (1989) *J. Biol. Chem. 264*, 3624–3628.

Lindqvist, Y., Brändén, C. I., Mathews, F. S. & Lederer, F. (1991) *J. Biol. Chem. 266*, 3198–3207.

Loh, E. V., Elliott, J. F., Cwirlo, S., Lanier, L. & Davis, M. M. (1989) *Science 243*, 217–220.

Mannervik, B. & Danielson, U. H. (1988) *CRC Crit. Rev. Biochem. 23*, 283–337.

McGroarty, E., Hsieh, B., Wied, D. M., Gee, R. & Tolbert, N. E. (1974) *Arch. Biochem. Biophys. 161*, 194–210.

Miles, C. S., Rouvière-Fourmy, N., Lederer, F., Mathews, F. S., Reid, G. A., Black, M. T. & Chapman, S. K. (1992) *Biochem. J. 285*, 187–192.

Monteilhet, C. & Risler, J. L. (1970) *Eur. J. Biochem. 12*, 165–169.

Nagai, K. & Thøgersen, H. C. (1984) *Nature 309*, 810–812.

Nakano, M. & Danowski, T. S. (1966) *J. Biol. Chem. 241*, 2075–2083.

Phillips, D. R., Duley, J. A., Fennell, D. J. & Holmes, R. S. (1976) *Biochim. Biophys. Acta 427*, 679–687.

Pompon, D., Iwatsubo, M. & Lederer, F. (1980) *Eur. J. Biochem. 104*, 479–488.

Reid, G. A., White, S., Black, M. T., Lederer, F., Mathews, F. S. & Chapman, S. K. (1988) *Eur. J. Biochem. 178*, 329–333.

Risler, Y., Tegoni, M. & Gervais, M. (1989) *Nucleic Acids Res. 17*, 8381.

Robinson, J. C., Keay, L., Molinari, R. & Sizen, I. W. (1962) *J. Biol. Chem. 237*, 2001–2010.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning, a laboratory manual*, 2nd edition, Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA 74*, 5463–5467.

Smith, D. B. & Johnson, K. S. (1988) *Gene (Amst.) 67*, 31–40.

Sullivan, P. A., Choong, Y. S., Schreurs, W. J., Cutfield, J. F. & Shepherd, M. G. (1977) *Biochem. J. 165*, 375–383.

Tiu, W. U., Davern, K. M., Wright, M. D., Board, P. G. & Mitchell, G. F. (1988) *Parasite Immunol. 10*, 693–706.

Tolbert, N. E. (1981) *Annu. Rev. Biochem. 50*, 133–157.

Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Buechter, D. D., Babbitt, P. C. & Kenyon, G. L. (1990) *Biochemistry 29*, 9856–9862.

Urban, P. & Lederer, F. (1984) *Eur. J. Biochem. 144*, 345–351.

Urban, P. & Lederer, F. (1985) *J. Biol. Chem. 260*, 11115–11122.

Urban, P., Chirat, I. & Lederer, F. (1988) *Biochemistry 27*, 7365–7371.

Volokita, M. & Somerville, C. R. (1987) *J. Biol. Chem. 262*, 15825–15828.

Xia, Z.-X. & Mathews, F. S. (1990) *J. Mol. Biol. 212*, 837–863.

Zaar, K. & Fahimi, H. D. (1991) *J. Histochem. Cytochem. 39*, 801–808.

Supplement material to:

Molecular cloning and nucleotide sequence of cDNA encoding rat kidney long-chain L-2-hydroxy acid oxidase

Expression of the catalytically active recombinant protein as a chimaera

Ahmed BELMOUDEN, K. H. Diêp LÊ, Florence LEDERER and Henri-Jean GARCHON

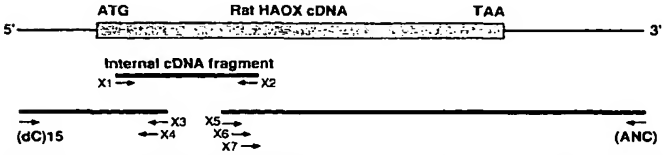


Fig. S1. Strategy for amplifying and cloning the cDNA of rat kidney hydroxy acid oxidase. The exact position in the final sequence and the sequence of the synthetic oligonucleotides X1–X7 and ANC are given in Table S1. The sequences of X3 and X4 and of X5 to X7 do not overlap (see Table S1) although the arrows symbolizing these primers are overlapping.

Table S1. Sequence and position of the primers used in the PCR amplification steps. X1, X2 and X7 are degenerate oligonucleotides.

Primer	Nucleotide position	Sequence
X1	64– 80	TGGGATTTTATTGAAGG C C C G A
X2	425–403	ACCATTTGTTTATTAAAATCCCA C C G G G
X3	168–149	TGACATATCTCTCAGGTATC
X4	108–189	ACTGTAGGTGATGCCGTCGT
X5	338–358	TGGAAGATATTGTTGCTGCTG
X6	374–392	GTTGGTTCCAACTCTACAT
X7	396–416	GTCAGACTGGGATTTCAACAA C T T
ANC		GACTCGAGTCGACATCG

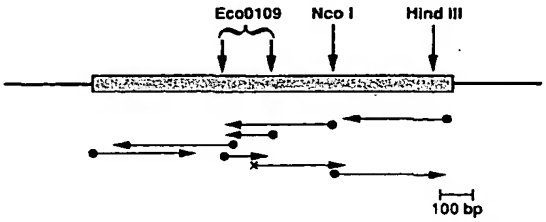


Fig. S3. Restriction map and sequencing strategy for the 3′-end of HAO cDNA (lower mass 1.1-kb fragment amplified from ss cDNA with X7 and ANC, see text). Restriction enzymes and their sites used to generate small fragments are shown above the thick bar representing the amplified segment. The lines below the bar represent the sequenced length of fragments obtained after subcloning (●) or after exonuclease-III deletion (×) (Henikoff, 1984). The higher-mass 1.1-kb fragment described in the text was sequenced entirely using the exonuclease-III deletion method. Arrows represent the direction of the sequencing.

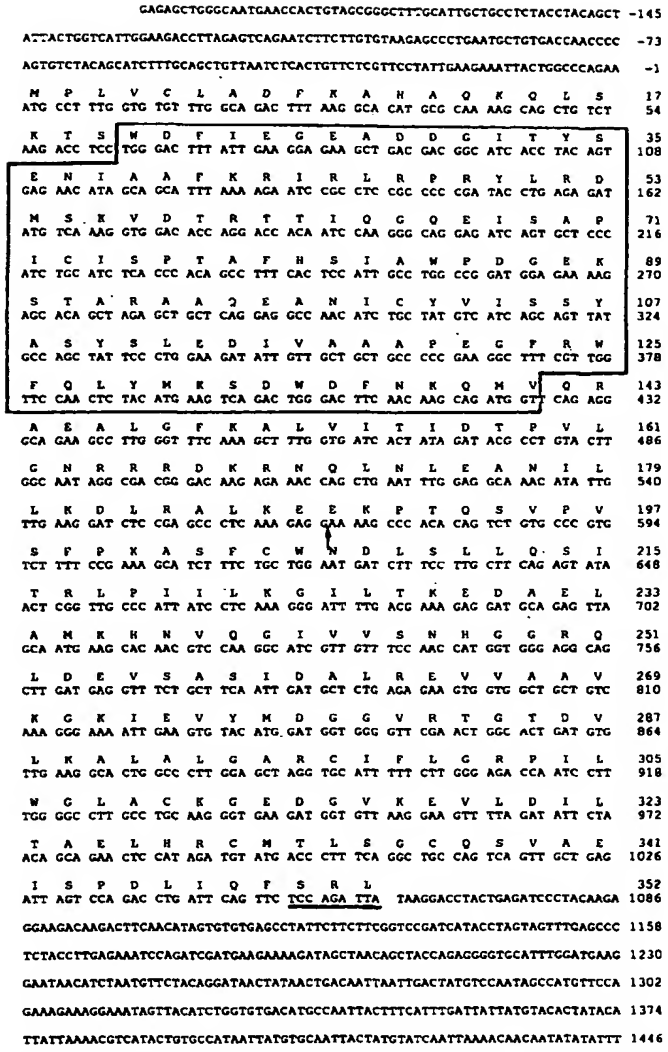


Fig. S2. Composite nucleotide sequence, and deduced amino acid sequence, of the cDNA of α-hydroxy acid oxidase from rat kidney. The boxed sequence is that obtained after the first PCR amplification with X1 and X2 (Table S1). The peroxisomal targeting sequence is underlined. The arrow indicates the position of the nine nucleotide insertion (TGAGAAAGG) found in some clones (see text).

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 268, No. 32, Issue of November 15, pp. 24330-24338, 1993
Printed in U.S.A.

The Cystic Fibrosis Transmembrane Conductance Regulator

OVEREXPRESSION, PURIFICATION, AND CHARACTERIZATION OF WILD TYPE AND ΔF508 MUTANT FORMS OF THE FIRST NUCLEOTIDE BINDING FOLD IN FUSION WITH THE MALTOSE-BINDING PROTEIN*

(Received for publication, May 17, 1993)

Young Hee Ko‡, Philip J. Thomas, Michael R. Delannoy, and Peter L. Pedersen§

From the Departments of Biological Chemistry and Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

The first nucleotide binding fold (NBF1) of the cystic fibrosis transmembrane conductance regulator (CFTR) and its disease-causing mutant form (ΔF508,NBF1) were overexpressed in high yield in *Escherichia coli* in fusion with the maltose-binding protein (MBP). The rationale for producing the chimerae was to aid in domain purification, solubilization, and crystallization and to examine the effect of protein-protein interactions on the properties of the mutant NBF1.

Both the purified wild type and ΔF508 mutant fusion proteins fold into functional nucleotide binding domains as determined by using the fluorescent nucleotide analog TNP-ATP (2′-(3′)-*O*-(2,4,6-trinitrophenyl)adenosine-5′-triphosphate). Moreover, the prominent secondary structural features of the two proteins as assessed by ultraviolet circular dichroism spectropolarimetry are very similar, as is the higher order structure evident in three separate protease digestion patterns. Finally, the stability of the nucleotide binding function of the two proteins is similar as assessed by sensitivity to urea.

Gel filtration chromatography and electron and confocal microscopy reveal that both fusion proteins, but not MBP alone, form organized fibers, suggesting that NBF1 self-associates, thus raising the possibility that CFTR may be oligomeric in the plasma membrane. Significantly, in the presence of high salt, these fusion proteins also have a propensity to form microcrystals. Finally, the two separate domains (NBF1 and MBP) constituting the fusion proteins appear to interact quite strongly as both proteins remain associated even after cleavage of their fusion junction.

The possible relevance of these novel findings to those approaches that might be taken to elucidate the three-dimensional structural differences between the wild type and ΔF508 mutant forms of CFTR, as well as to ameliorate the severity of cystic fibrosis, is discussed.

The most common lethal autosomal recessive genetic disease in the Caucasian population is cystic fibrosis (CF).[1] The disease results from mutations in the gene that codes for the cystic fibrosis transmembrane conductance regulator (CFTR) protein comprising 1480 amino acid residues (1–3). Predictions suggest that CFTR embraces two transmembrane domains, each having six putative membrane-spanning segments; two nucleotide binding folds, NBF1 and NBF2, each possessing a Walker A and B consensus sequence (4); and a regulatory (R) domain having numerous charged amino acid residues, which can be phosphorylated by some cAMP-dependent protein kinases (5, 6).

The majority of the CFTR mutations occur within or near the two nucleotide binding folds. Significantly, 92% of all patients with cystic fibrosis have at least one allele in which the codon for phenylalanine 508 in NBF1 is deleted (7–9). Strong evidence has established CFTR as a chloride ion channel regulated by cAMP-dependent phosphorylations (10–16). Moreover, recent studies indicate that interaction of intracellular ATP with NBF1 (or NBF2) may be essential for chloride channel activation (17, 18). Hence, the roles of the NBFs may be of great importance in elucidating the molecular basis of CF.

A synthetic peptide embracing the Walker A consensus sequence and the region surrounding Phe-508 contains less β-sheet structure in the ΔF508 mutant than in the wild type peptide (19, 20). Significantly, the peptide studies demonstrated also that this structural change is due to destabilization of the mutant peptide structure, indicating that the ΔF508 mutation results in a protein folding defect (19–21). Recent *in vivo* studies support this finding, as lowering the temperature of cells harboring the ΔF508 mutant protein increases the amount of functional protein at the plasma membrane (22, 23). Despite the importance of this observation, the functional mutant protein exhibits reduced open channel probabilities relative to the wild type protein (16, 22). The structural basis of the altered channel function of the mutant protein in the membrane is currently unknown. Therefore, the production of both purified wild type and ΔF508 mutant forms of NBF1 in a state amenable to biochemical and structural studies is necessary to define the

* This work was supported in part by National Institutes of Health Grant DK 43962 and a grant from the Cystic Fibrosis Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of a postdoctoral fellowship from the Johns Hopkins University Institute for Biophysical Research on Macromolecular Assemblies and of a New Investigator Grant from the Cystic Fibrosis Foundation.

§ To whom correspondence should be addressed: Dept. of Biological Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205.

[1] The abbreviations used are: CF, cystic fibrosis; CFTR, cystic fibrosis transmembrane conductance regulator; NBF, nucleotide binding fold; TNP-ATP, 2′-(3′)-*O*-(2,4,6-trinitrophenyl)adenosine-5′-triphosphate; HPLC, high performance liquid chromatography; PAGE, polyacrylamide gel electrophoresis; X-gal, 5-bromo-4-chloro-3-indolyl-β-galactoside; ΔF508, deletion of phenylalanine 508 from CFTR; MBP, maltose-binding protein; PCR, polymerase chain reaction; MDR, multidrug-resistant P-glycoprotein.

structural change in the mutant protein that leads to altered function.

With the above thoughts in mind, we rationalized that perhaps fusion of NBF1 and its ΔF508 mutant form with a highly soluble protein might facilitate purification and structural characterization. Precedence for this approach derives from the recent successful elucidation of the three-dimensional structure of actin by first forming a complex with DNase 1 (24). In the studies reported in detail below, we chose to fuse both the wild type and ΔF508 mutant NBF1 domains with the maltose-binding protein, a soluble protein for which a high resolution crystal structure has already been determined (25).

EXPERIMENTAL PROCEDURES

Materials

Clones T16-1 and Cl-1/5 were obtained from American Type Culture Collection (ATCC). Primers used in the PCR procedures were synthesized in the Johns Hopkins University Protein/Peptide/DNA Synthesis Facility using an Applied Biosystems model 380B synthesizer. The pMAL-cR1 expression vector, the purification system, Factor Xa, restriction enzymes, and T4 DNA ligase were obtained from New England Bio-Labs. TNP nucleotides were obtained from Molecular Probes. *Pfu* DNA polymerase and the Plasmid Quick DNA Purification Kit were purchased from Stratagene. The DNA sequencing kit (Sequenase version 2.0 DNA) was from U. S. Biochemical Corp., and the maltose-binding protein (MBP) from New England Bio-Labs. Waters-Millipore supplied the Protein Pak 300 SW HPLC gel filtration column (7.5 mm inner diameter × 30 cm length), and uranyl acetate was purchased from Ted Pella. All other reagents were of the highest quality commercially obtainable.

Methods

Construction of the Expression Vector Containing CFTR NBF1—An outline of the approach used is summarized in Fig. 1. The expression cassette polymerase chain reaction (EC-PCR) (26, 27) was employed to synthesize the NBF1 of CFTR. Restriction enzyme sites (*Bgl*II and *Hind*III), GC clamps, and a stop codon were incorporated into the primers, including a *Bgl*II site in the forward primer and a *Hind*III site in the reverse primer. The sequence of the primers used to define and amplify NBF1 from Phe-433 (base 1429) to Ser-589 (base 1899) are as follows.

Forward primer: 5′-CGCGCAGATCTTTCTCACTTCT
(35 bases) TGGTACTCCTGTC-3′

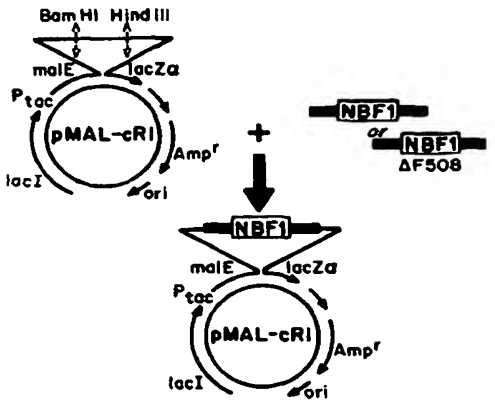


FIG. 1. **Construction of the expression vector containing NBF1 of CFTR.** The cDNA of NBF1 or ΔF508,NBF1 is ligated into *Bam*HI/*Hind*III sites of pMAL-cR1 in frame with the MBP (*malE* gene product). The insertion of the NBF1 cDNA into this plasmid results in disruption of the *lacZα* gene and allows blue to white colony selection when bacterial cells are grown on plates containing IPTG and X-gal. The pMAL-cR1 is 6133 base pairs long and the NBF1 and its mutant are, respectively, 471 and 468 base pairs long. For experimental details, see "Methods."

Reverse primer: 5′-GCGCAAGCTTTTAGCTTTCAA
(36 bases) ATATTTCTTTTTCTG-3′

The expression cassette PCR was carried out in a 100-μl reaction volume composed of 20 mM Tris-Cl, pH 8.8, 2.5 units of *Pfu* polymerase, 0.5 μM primers, 50 ng of template cDNA, 200 μM dNTPs, 1.5 mM $MgCl_2$, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, and 0.1% Triton X-100. The conditions for each PCR cycle was as follows: denaturation at 94 °C for 43 s, annealing at 43 °C for 1 min, and elongation at 73 °C for 1 min and 15 s.

Clones T16-1 and Cl-1/5 were used as templates to amplify the wild type NBF1 and the mutant ΔF508,NBF1, respectively. These template plasmids for the PCR were prepared by use of Stratagene's Plasmid Quick Midi Column. PCR amplified products (~468–471 base pairs expected) were purified by use of Bio-Rad's Prep-A-Gene DNA Purification Matrix kit and were subsequently digested with *Bgl*II and *Hind*III restriction enzymes. The digested PCR products were then subjected to agarose gel electrophoresis, further purified by use of Bio-Rad's Prep-A-Gene DNA Purification Matrix kit, and were finally ligated into the *Bam*HI/*Hind*III-cut pMAL-cR1 expression vector (28, 29) in frame with the MBP. The DNA sequence encoding for the protein sequence -I-E-G-R-, which is recognized by the protease called Factor Xa, is present between the DNA sequences for MBP and NBF1. The insertion of NBF1 into the vector disrupts the expression of *lacZα* DNA and provides a rapid selection method for the correct recombinant. Thus, white colonies instead of blue colonies are formed on X-gal in an α-complementing host such as TB1 (30) or JM107 (31).

Competent TB1 *Escherichia coli* were transformed with the above ligation mixture and grown on LB/agar/ampicillin plates. Plasmids purified from the ampicillin-resistant white colonies were sequenced by Sanger's method (32, 33) to confirm the fidelity of the PCR step and formation of the proper construct. All other molecular biological methods employed were conducted by standard procedures (34).

Overexpression and Purification of MBP-NBF1 Fusion Proteins—TB1 *E. coli* harboring the appropriate recombinant plasmids were grown at 37 °C in LB media containing 0.2% glucose and ampicillin (0.1 mg/ml) until the absorbance at 600 nm reached about 0.5 units (~2×10^{10} cells/ml). Cells were then induced with 0.3 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Incubation was continued for an additional 3 h after which the cells were harvested by centrifugation at 4000 × *g* for 20 min. A combination of lysozyme and sodium deoxycholate was used to lyse the cells (35, 36). The lysate was centrifuged to remove the broken cell walls and debris (Beckman Ultracentrifuge L7-55; SW 28 rotor; 4 °C, 28,000 rpm, for 1 h). The supernatant was then incubated with 55% saturated ammonium sulfate at 4 °C for 30 min, and centrifuged at 12,000 × *g* for 40 min at 4 °C in a Sorvall model RC 2B centrifuge equipped with a SS 34 rotor (radius = 4.25 inches). The pellet was resuspended in 40 ml of the column buffer (10 mM sodium phosphate, 0.5 M NaCl, and 1.0 mM EGTA, pH 7.2) to give a protein concentration of ~2.5 mg/ml.

Purification was carried out by loading the resuspended sample onto a column containing amylose resin that had been previously equilibrated with the column buffer. The binding capacity of the column is 2 mg of fusion protein/ml of bed volume. The column was washed with three column volumes of column buffer containing 0.25% Tween 20 and finally with column buffer without the surfactant. Elution of the bound fusion protein was accomplished at a flow rate of 1 ml/min with 20 mM maltose dissolved in the column buffer. Fractions containing the fusion proteins, detected by UV absorbance at 280 nm or by Coomassie dye binding, were pooled and concentrated using Amicon's Centri-Prep 30 concentrator.

SDS-PAGE was carried out as described by Laemmli (37). Gels were stained with Coomassie dye to determine the degree of overexpression of the recombinant protein and to estimate its purity. Gas phase sequencing of the N terminus (38, 39) of the purified protein, as well as an amino acid analysis, established that the product isolated was the desired fusion protein.

TNP-ATP Binding—Fluorescence enhancement of the ATP analog TNP-ATP (40–42) was used to study the nucleotide binding characteristics of the fusion proteins in the presence or absence of urea. The procedure used, which includes correcting for the inner filter effect on the fluorescence of the TNP-ATP probe, has been described previously (19). Samples (4 μM protein in 20 mM HEPES, pH 7.4, at concentrations of TNP-ATP indicated) were excited at 410 nm (slit width, 1 mm). The emission was then measured at 550 nm (slit width, 2 mm) using an Aminco model SPF-125 spectrofluorometer. Finally, ATP was added after TNP-ATP binding to dem-

onstrate that ATP could displace bound TNP-ATP.

Protease Susceptibility—To the purified fusion proteins at concentrations of 60–70 μM in a 20 mM HEPES buffer, pH 7.4, proteases (trypsin, endoproteinase Glu-C, or endoprotease Arg-C) were added at 1% (w/w) in the presence or absence of 12.5 mM ATP plus 42 mM maltose. The mixtures were incubated at 23 or 37 °C for times up to 90 min. Aliquots of the digests were taken at varying times, subjected to SDS-PAGE, and then visualized after staining with Coomassie Blue dye.

Circular Dichroism Spectroscopy (CD)—To assess the secondary structures of the fusion proteins, CD spectra were obtained at 23.2 °C in a 2-mm path length cuvette using an AVIV spectropolarimeter. Wild type and mutant fusion proteins were present at a concentration of 1.3 μM in a 0.4-ml system containing 10 mM Tris-Cl buffer, pH 7.6. Details of the procedure have been published previously (20).

Gel Filtration HPLC—Fusion proteins (175 μg in 1 ml) were loaded onto a Protein Pak 300 SW filtration column that had been equilibrated previously with 0.2 M HEPES at pH 7.4. Elution profiles were monitored at 280 nm by an isocratic gradient using a flow rate of 0.5 ml/min. Molecular size markers used for gel filtration chromatography were blue dextran, β-amylase, alcohol dehydrogenase, bovine serum albumin, and MBP, corresponding to 2000, 200, 150, 66, and 42 kDa, respectively.

Amino Acid and N-terminal Sequence Analysis—For amino acid analysis, the fusion protein (1280 pmol of the wild type or 920 pmol of the ΔF508 mutant) was vapor-phase-hydrolyzed in 6 N HCl at 110 °C for 24 h. The resulting hydrolyzed amino acid residues were derivatized with phenyl isothiocyanate, and the derivatives were separated by reverse phase chromatography and subsequently analyzed by PICO-TAG amino acid analysis system software.

For the N-terminal sequencing of the fusion protein, 100–120 pmol were electroblotted from an SDS gel onto a polyvinylidene difluoride membrane. The putative 60-kDa molecular mass was excised and subjected to Edman degradation chemistry (38, 39) using an Applied Biosystems model 475A protein sequencing system, which employed both a gas and pulsed liquid phase.

Electron Microscopy—Parloidion-coated 200-mesh copper grids were stabilized with carbon and glow discharged before applying 1–2 μM of the fusion proteins in 20 mM HEPES buffer at pH 7.4. The protein samples in a 30-μl volume were adsorbed onto the pretreated grid for 30 s. The grid was then washed successively three times in drops of 20 mM HEPES. Finally, the treated grid was floated on a drop of filtered 1% aqueous uranyl acetate for 1 min. The excess solution was blotted with Whatman no. 2 filter paper. The negatively stained sample on the grid was then examined with a Zeiss transmission electron microscope (model 10A) at an acceleration voltage of 60 kV. The magnification was calibrated using an appropriate diffraction grating replica. For a qualitative determination of the dimensions of the protein fibers, micrographs were printed at a magnification of 36,000 or 98,000. A Photomicrograph Scale Marker™ was used to draw scale marks on the print of each electron micrograph.

Confocal Microscopy—Confocal microscopy of the fusion proteins (5 μM) in a volume of 0.5 ml in 20 mM HEPES, pH 7.4, was performed at 23 °C in the presence of 20 μM TNP-ATP. Following an incubation of 10 min, dialysis of the sample against 10 mM Tris-Cl at pH 8.0 (two changes with 300 ml at 35-min intervals) was conducted to remove the unbound TNP-ATP. The samples were added to an equal volume of 0.1% *p*-phenylenediamine in a 1:9 (v/v) phosphate-buffered saline solution in glycerol. Cover glasses were placed over specimens of the fusion protein samples on microscope slides and the edges sealed with a cellulose acetate polymer solution to make them ready for examination with a Bio-Rad MRC-600 laser scanning confocal imaging system. Images of the emission fluorescence at 520 nm of the TNP-ATP that was bound to the protein were obtained. For the three-dimensional computer-assisted images, two identical sets of optical sections were projected in the *z* axis, pixel-shifted, and superimposed to create the resulting stereo images.

Crystallization—Microcrystals were obtained using vapor diffusion through hanging drops (43). Fusion proteins at a concentration of 3 mg/ml in a volume of 48 μl were incubated with 42 μM TNP-ATP for 30 min at 23 °C. Following this incubation, a 2-μl aliquot from the above reaction mixture was combined with 2 μl of the precipitating solution containing 0.2 M lithium sulfate, 30% polyethylene glycol 4000, and 0.1 M Tris-Cl, pH 9.0.

Cleavage of the MBP-NBF1 Fusion Protein—The MBP-NBF1 fusion protein was treated with 0.08% SDS, incubated at 23 °C for 16 h, and then subjected to the action of Factor Xa at a protease to fusion protein ratio of 1:100 for 14 h at 4 °C. The degree of cleavage was monitored by SDS-PAGE.

Determination of Protein—Protein was estimated either by the method of Bradford (44) using the Coomassie protein assay reagent kit from Pierce, or where indicated from amino acid compositional analysis.

RESULTS

Overexpression and Purification of MBP-NBF1 and MBP-(ΔF508)NBF1—The correct orientation and cDNA sequence of CFTR nucleotide binding fold 1 and its ΔF508 mutant in the pMAL-cR1 expression vector in frame with the MBP was first confirmed by sequencing the recombinant plasmids (see "Methods"). The entire codon for phenylalanine in the mutant was absent. As shown in Fig. 2, the overexpression of fusion proteins under the control of the *tac* promoter after induction with IPTG is strikingly evident when the induced expression level is compared with the uninduced cell lysate. As the MBP exhibits a high affinity for amylose, the purification of the fusion proteins by amylose column affinity chromatography was simple and rapid. A yield of approximately 25 mg of purified proteins per liter of cell culture was achieved. The resultant proteins appear >95% pure as judged by SDS-PAGE (Fig. 2) and have the expected apparent molecular masses of 60 kDa (42 kDa for MBP + 18 kDa for NBF1). Moreover, amino acid analysis gave experimental to theoretical ratios very close to 1 in most cases (Table I). Significantly, the ratio of picomoles of phenylalanine to the number of picomoles of other amino acids, obtained from amino acid analysis, is consistently lower for the ΔF508 mutant fusion protein than for the wild type, confirming that the number of phenylalanine residues is less in the mutant. Finally, in experiments not presented here, N-terminal se-

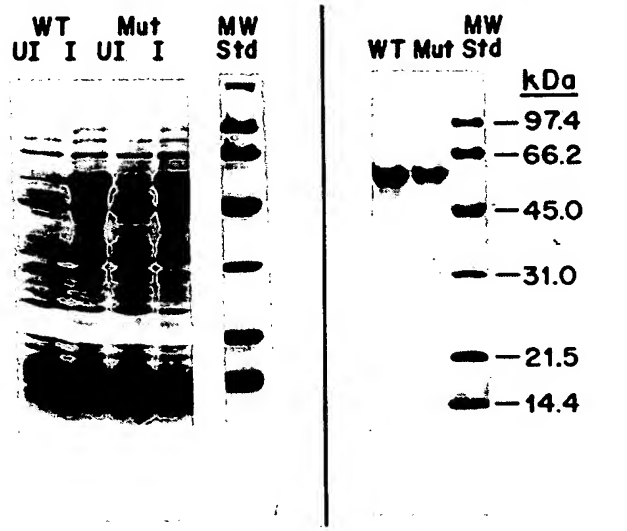


FIG. 2. **Analysis by SDS-PAGE of overexpressed and purified wild type and mutant MBP-NBF1 fusion proteins.** *Left panel*, to demonstrate the overexpression of the MBP-NBF1 fusion proteins the whole cell lysate (300 μl, 2 × 10^8 cells/ml) before IPTG induction was solubilized in SDS sample buffer and loaded onto the gel designated as "uninduced cells" (*UI*). SDS-PAGE was then carried out by the method of Laemmli (34). The *E. coli* cell lysate (75 μl, stationary phase cells) after IPTG induction was treated in the same manner and designated "induced cells" (*I*). Molecular size markers are, from *top* to *bottom*, phosphorylase *b* (97.4 kDa), bovine albumin (66.2 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), trypsin inhibitor (21.5 kDa), and lysozyme (14.4 kDa). *Right panel*, the fusion proteins (25 μg) purified by amylose affinity chromatography were subjected to SDS-PAGE. Both proteins exhibit an approximate molecular mass of 60 kDa.

quence analysis (10 amino acids) also confirmed the identity and purity of the fusion proteins.

TNP-ATP Binding Characteristics—TNP-ATP was found to bind similarly to both the wild type and mutant fusion proteins as depicted in Fig. 3. ATP effectively displaces all of the bound TNP-ATP, indicating that the site involved is capable of binding the physiological substrate. The apparent K_D values determined according to a method previously reported (45) for binding of TNP-ATP to the wild type and mutant fusion proteins are 1.8 and 2.3 μM, respectively. The corrected K_D values for ATP binding, determined by competition analysis between TNP-ATP and ATP, are 1.8 and 1.9 mM, respectively. The stoichiometry of TNP-ATP binding is approximately 1 for both types of fusion proteins under the conditions employed. In the control experiment, MBP alone did not bind TNP-ATP even at concentrations exceeding 5 μM. This important control emphasizes that the nucleotide binding properties of the fusion proteins, as expected, are due to interaction of TNP-ATP or ATP with NBF1 or ΔF508,NBF1 *per se* and not with the MBP.

TABLE I

Amino acid analysis of CFTR wild type and ΔF508 MBP-NBF1 fusion proteins

Analyses were carried out as described under "Methods" on fusion proteins that had been overexpressed in *E. coli* and purified by amylose column affinity chromatography, also as described under "Methods." The mol ratios of the calculated amino acid residues to the theoretical values are based on 48 mol of Leu/mol of fusion protein. A yield of 100% was assumed for the leucine residues.

Amino acid[a] (*X*)	Wild type (experimental/ theoretical)	Mol ratio (Phe/*X*)	Mutant (experimental/ theoretical)	Mol ratio (Phe/*X*)
Ala	1.02	0.448	1.01	0.433
Arg	1.20	1.444	1.20	1.382
Asx[b]	1.16	0.366	1.15	0.356
Cys	0.85	14.280	0.83	14.006
Gly	0.96	0.536	0.96	0.515
Glx[c]	1.12	0.393	1.12	0.379
His	1.63	3.735	1.62	3.588
Ile	0.84	0.739	0.85	0.701
Leu	1.00	0.506	1.00	0.484
Lys	1.06	0.485	1.06	0.467
Met	1.33	1.657	1.37	1.547
Phe	1.01	1.000	1.01	1.000
Pro	0.94	0.994	0.93	0.962
Ser	0.99	0.724	0.99	0.691
Thr	0.92	0.980	0.89	0.966
Tyr	1.01	1.155	1.03	1.076
Val	0.98	0.850	1.02	0.783

[a] Tryptophan is normally degraded and not detected.
[b] Asx, aspartic acid + asparagine.
[c] Glx, glutamic acid + glutamine.

Secondary Structure as Assessed by Circular Dichroism Spectroscopy—The overall secondary structures of the wild type and ΔF508 mutant fusion proteins were estimated by CD spectropolarimetry. As shown in Fig. 4, the resultant spectra, which summarizes the mean residue molar ellipticity as a function of wavelength, reveals no obvious differences between the wild type and ΔF508 mutant fusion proteins. In both cases, the mean residue molar ellipticity of the fusion proteins exhibits a maximum at 197 nm and two minima, one at 209 nm and the other at 223 nm. Deconvolution of the spectra using the Prosec program (46) predicts that both wild type and ΔF508 mutant fusion proteins contain nearly identical secondary structural elements distributed as approximately 40% α-helix, 24% β-strand, 13% turns, and 24% random coil. Thus, both fusion proteins are highly structured as assessed by this technique.

Protease Susceptibility—To test for subtle conformational differences between the wild type and the ΔF508 mutant, both fusion proteins were incubated with three different proteases (trypsin, endoproteinase Glu-C, and endoproteinase Arg-C), and their susceptibility to proteolysis was analyzed by 10% SDS-PAGE. Trypsin in the absence (Fig. 5*A*), and in the presence (Fig. 5*B*) of 12.5 mM ATP plus 42 mM maltose produces essentially identical digestion patterns for the two fusion proteins. ATP plus maltose seems to have no effect in protecting the proteins from trypsin digestion (Fig. 5*B*). Furthermore, the kinetics of the degradation of the two fusion proteins by trypsin is identical. Trypsin was the most effective in 3-min digestion, where the starting proteins were mostly degraded. Endoproteinase Glu-C required about 20 min to degrade the fusion proteins to a similar degree (Fig. 5*C*). Endoproteinase Arg-C was found to be the least effective (Fig. 5*D*). Like that of trypsin, the digestion patterns of the endoproteinases Glu-C and Arg-C are very similar for both wild type and mutant fusion proteins.

Susceptibility to Denaturation by Urea—Previous studies with synthetic peptides of NBF1 have shown that loss of the TNP-ATP binding function in the presence of the denaturant urea correlates with loss of secondary structure (19, 20). This simple test has been used here to assess the relative stabilities of the mutant and ΔF508 fusion proteins. As shown in Fig. 6

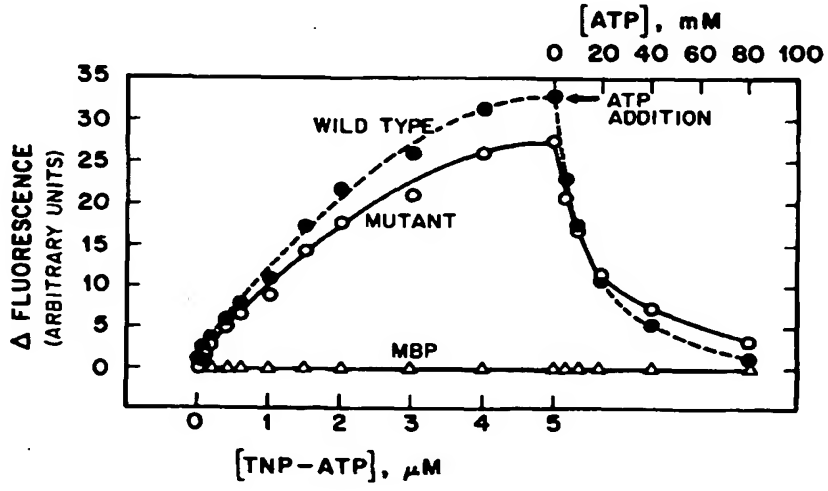


FIG. 3. **The nucleotide binding function of MBP-NBF1 fusion proteins.** The enhanced extrinsic fluorescence of TNP-ATP upon binding to the fusion proteins was observed at 550 nm emission wavelength after exciting the samples at 410 nm. When ATP was added to the above sample to compete off the bound TNP-ATP, the fluorescence of the TNP-ATP rapidly decreased. See "Methods" for details.

(*A* and *B*), both fusion proteins lose more than half their capacity to bind TNP-ATP at a concentration of 1.8 M urea, and at 3.6 M urea most of the binding capacity is lost in both cases. Thus, this test indicates that the stability of the nucleotide binding function of the wild type and ΔF508 mutant fusion proteins are similar.

Formation of Highly Organized Macromolecular Species—The tendency of both the wild type and ΔF508 fusion proteins to form high molecular weight species is evident from the chromatograms of gel filtration experiments (Fig. 7, *A* and *B*). Significantly, both fusion proteins elute at the void volume of an HPLC gel filtration column, exhibiting a molecular exclusion limit > 2000 kDa as defined by blue dextran (Fig. 7*C*). In contrast, the MBP, which has a molecular mass of 42 kDa, elutes as one major and two minor peaks all well within the inclusion volume of the column.

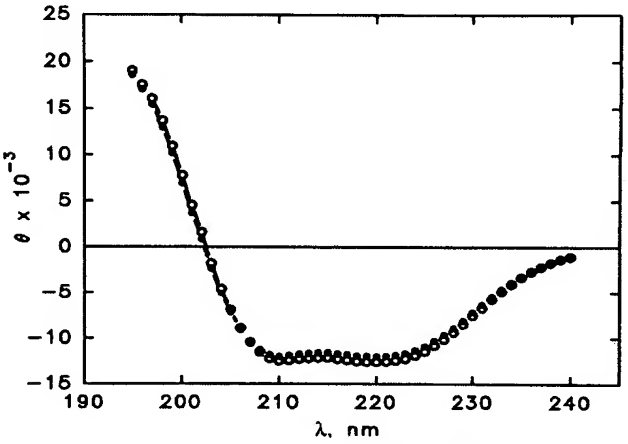


FIG. 4. **Circular dichroism spectra of the MBP-NBF1 fusion proteins.** The molar mean residue ellipticity of the wild type or the mutant protein as a function of wavelength is depicted. Fusion proteins (1.3 μM) both dissolved in 10 mM Tris-Cl, pH 7.6, were subjected to CD spectroscopic studies using a 2-mm path length cuvette as described under "Methods."

Electron microscopy studies with negatively stained samples indicate that both the wild type and ΔF508 mutant fusion proteins are present in the form of long, highly organized fibers exhibiting widths of about 20 nm, and that individual fibers tend to associate to form bundles having collective widths of 50–380 nm (Fig. 8*A*). As shown by confocal fluorescence imaging (Fig. 8, *B* and *C*), TNP-ATP can bind throughout these organized fibrous networks, indicating that each fusion protein within the network has retained the capacity to bind nucleotide. The fiber bundles comprising fusion proteins to which TNP-ATP is bound are about 1500 nm in width.

Formation of Microcrystals—When fusion proteins in the presence of TNP-ATP were subjected to conditions commonly used to crystallize proteins (see "Methods"), the formation of microcrystals occurred in the presence of lithium sulfate and polyethylene glycol (Fig. 8*D*). Crystals (≤0.005 mm) were apparent usually after 3 weeks and have the light orange color characteristic of the TNP-ATP molecule. Microcrystals were not formed under identical conditions in the absence of the fusion proteins.

Proteolytic Cleavage of the MBP-NBF1 Fusion Protein Followed by Chromatography on an Amylose Column—The fusion junction between the MBP and NBF1 contains a recognition site for the protease involved in blood clotting called Factor Xa. Attempts to cleave this site using a variety of conditions in which ionic strength, various detergents, solvents, pH, and temperature were tested proved unsuccessful, suggesting a strong interaction between the MBP and NBF1 within the fusion protein that masked access of the protease to the cleavage site. For this reason, a low concentration of SDS was employed in the proteolytic incubation medium, and as shown in the subsequent SDS-PAGE profiles (Fig. 9*A*), NBF1 is now almost completely cleaved. However, if the cleaved fusion protein is first subjected to chromatography on an amylose column, the NBF1 remains associated with the MBP and coelutes with it upon addition of the maltose containing buffer

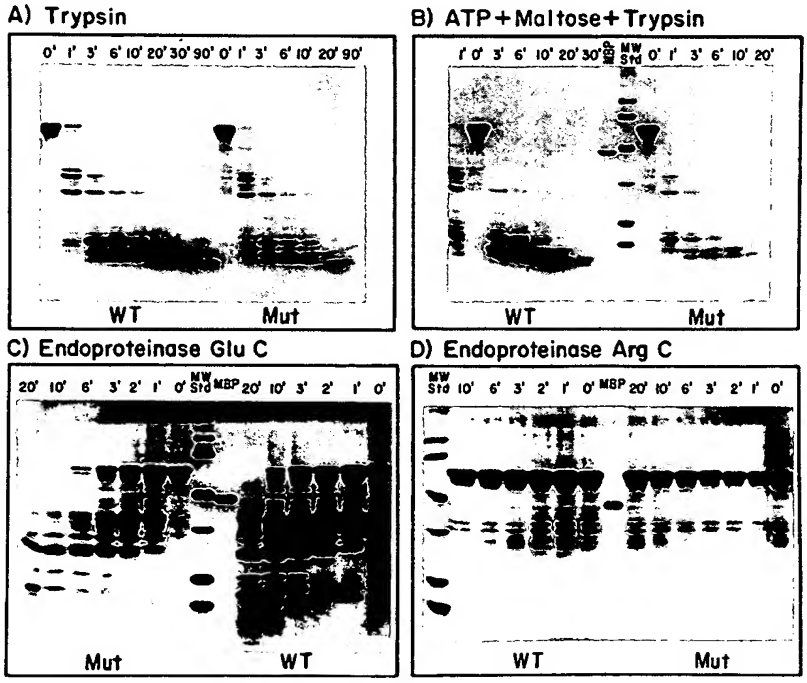


FIG. 5. **Susceptibility of MBP-NBF1 fusion proteins to proteases.** The digestion patterns of fusion proteins analyzed by SDS-PAGE are depicted. *A*, trypsin; *B*, trypsin in the presence of 42 mM maltose plus 12.5 mM ATP; *C*, endoproteinase Glu-C; *D*; endoproteinase Arg-C. Times of digestion are indicated. See "Methods" for details. Molecular size markers are identical to those noted in the legend to Fig. 2.

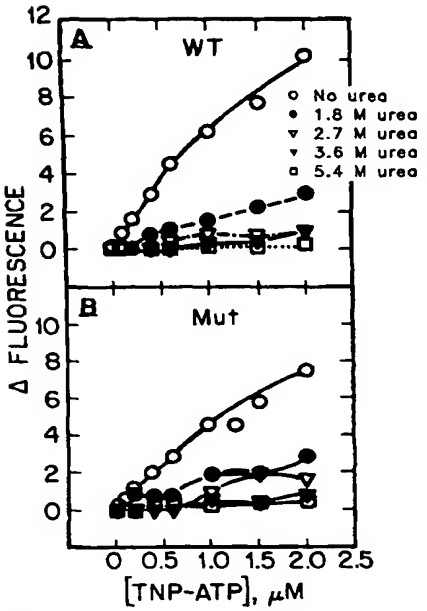


FIG. 6. **Effect of urea on the nucleotide binding function of fusion proteins.** The wavelengths for excitation and emission of TNP-ATP are the same as in Fig. 3. Emission fluorescence of TNP-ATP bound to the wild type (*A*) and the mutant (*B*) at varying concentrations of urea is shown.

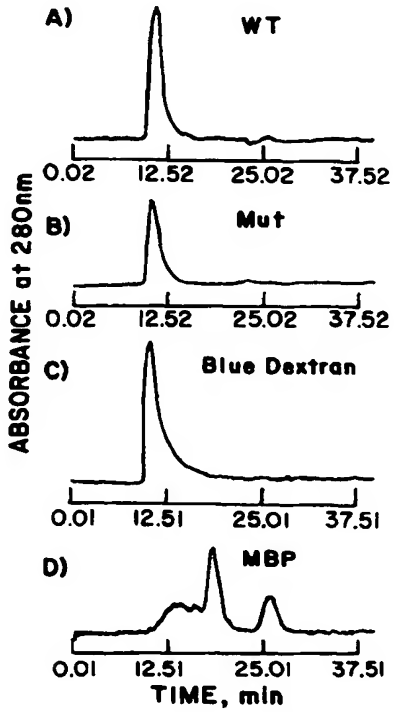


FIG. 7. **Gel filtration HPLC on the MBP-NBF1 fusion proteins.** HPLC chromatography was carried out exactly as described under "Methods." Retention times for wild type (*A*), mutant (*B*), and blue dextran (*C*) are 11 min as depicted in the elution profiles of gel filtration chromatogams. MBP (*D*) eluted as a major peak with a retention time of 16.5 min with two minor peaks with retention times, respectively, of 14.0 and 18.5 min.

(Fig. 9*B*). The identity of the cleaved NBF1 was confirmed by sequencing the N terminus of the protein band designated with the *lower arrow* in Fig. 9*B*. (It should be noted in Fig. 9 (*panel A*, *lane 3*, and *panel B*, *lanes 3* and *4*) that Factor Xa causes some nonspecific cleavage of the maltose-binding protein and the NBF1 as indicated by several additional bands.)

These results indicate that within the MBP-NBF1 fusion complex a tight association occurs between the two fusion partners that is retained even after cleavage of the fusion junction.

DISCUSSION

The novel experiments described here are part of a longer term effort directed at understanding the molecular and chemical basis of cystic fibrosis. In the vast majority of cases of the disease (~92%), the problem distills down to first elucidating those structural perturbations in the CFTR protein resulting from deletion of phenylalanine 508 within the first nucleotide binding fold (NBF1), and then elucidating how these perturbations impair ATP-dependent chloride channel activity. Previous experiments in this laboratory conducted with long synthetic peptides have demonstrated directly that NBF1 binds ATP (19). In addition, they indicate that Phe-508 lies within a β sheet region, that its deletion may result initially in a register shift to restore the β sheet (17), but that the resultant structure is less stable (20), and, depending on conditions, may unfold impairing ATP binding essential for normal chloride channel activity (20, 21, 47). Significantly, this simple working model is supported by recent studies in eukaryotic cells demonstrating that at low temperature some CFTR can reach the plasma membrane in functional form, albeit reduced, whereas at higher temperature (*e.g.* 37 °C) this apparently does not occur (22, 23).

Operating within the framework of the above model, the objective of this study was to obtain in large amounts purified preparations of wild type and ΔF508 mutant NBF1 domains in which structural changes are minimal. Assuming that the initial change resulting from the Phe-508 deletion is a register shift within a region of secondary structure, the initial perturbation may be quite subtle, detectable only by sophisticated structural approaches, *i.e.* x-ray crystallography or NMR spectroscopy. To this end, we were able to successfully fuse both the wild type and ΔF508 NBF1 domains to the maltose-binding protein, a protein for which a crystal structure is known at high resolution. Moreover, these fusion proteins could be overexpressed in large amounts and rapidly purified in a single step (Fig. 2).

The resultant wild type and ΔF508 mutant fusion proteins were shown to be virtually indistinguishable in their secondary structural properties as assessed by circular dichroism spectrosopy (Fig. 4), a finding supported by experiments showing that both fusion proteins exhibit similar proteolytic degradation profiles (Fig. 5). The finding that the resultant fusion proteins both bind amylose and TNP-ATP (Figs. 2 and 3) demonstrates that within the fusion complex both the MBP and NBF1 have folded correctly to allow formation of their respective substrate binding sites. Finally, the additional observation that the loss of TNP-ATP binding capacity as a function of urea concentration is similar (Fig. 6) indicates that within the fusion proteins no major differences in stability exist between wild type and ΔF508 mutant NBF1 domains under the conditions tested. As the maltose-binding protein remains tightly associated with NBF1 even after cleavage of the MBP-NBF1 fusion junction (Fig. 9), direct protein-protein interactions between the two fusion partners may be responsible for minimizing major differences between wild

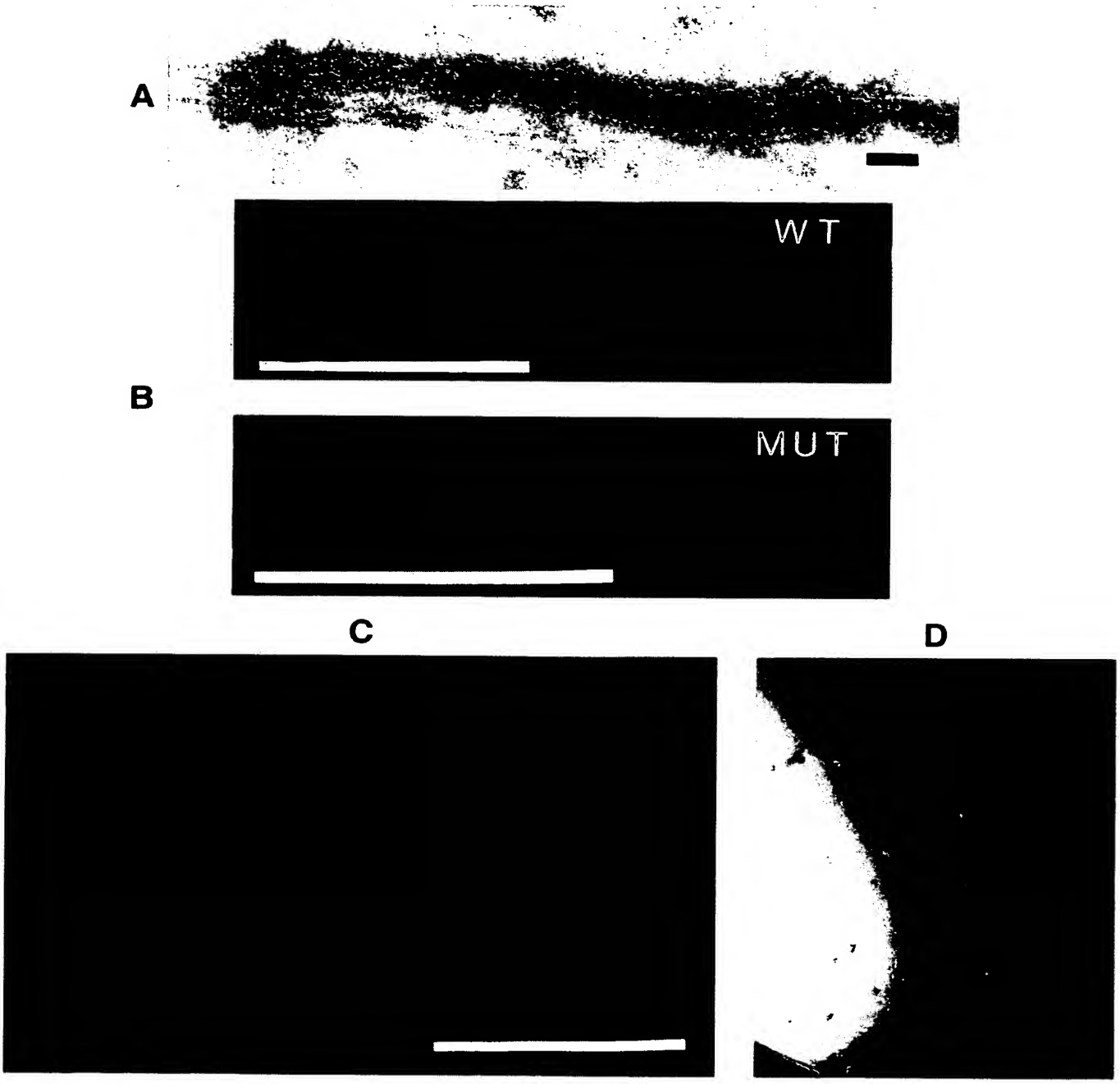


FIG. 8. **Electron micrographs, confocal fluorescence images, and microcrystals of MBP-NBF1 fusion proteins.** *A*, electron micrograph of wild type MBP-NBF1 fusion protein negatively stained with uranyl acetate (see "Methods") depict their fibrous structures. The calibration bar shown on *A* represents 100 nm in length; magnification, × 98,000. (Similar micrographs were obtained also for the ΔF508, MBP-NBF1 fusion protein.) *B*, confocal fluorescence microscopy was carried out exactly as described under "Methods," after which colors of the fibers were arbitrarily assigned. Emission fluorescence of TNP-ATP bound to the wild type (*blue*) and the mutant (*red*) at 520 nm are represented as two-dimensional images. Both are visualized in the form of fibrous structures. The length of the calibration bar represents 25 μm. *C*, confocal fluorescence microscopy of the wild type fusion protein depicting stereo images (superimposed *red/blue* anaglyph) of fibrous structures in three dimensions. *D*, microcrystals of the MBP-NBF1 fusion protein. Crystals were obtained by using vapor diffusion employing the hanging drop method (see "Methods").

type and ΔF508 mutant NBF1 domains.

The additional finding reported here that the MBP-NBF1 fusion proteins, and not the MBP alone, form highly organized macromolecular species in aqueous media (Fig. 8), is fascinating and of considerable interest for several reasons. First, there have been reports (48, 49) indicating that the related multidrug-resistant P-glycoprotein (MDR) may exist as a dimer or higher ordered oligomeric species in the plasma membrane. The same may be the case for CFTR. Within such a complex, stability may be maintained in part by direct

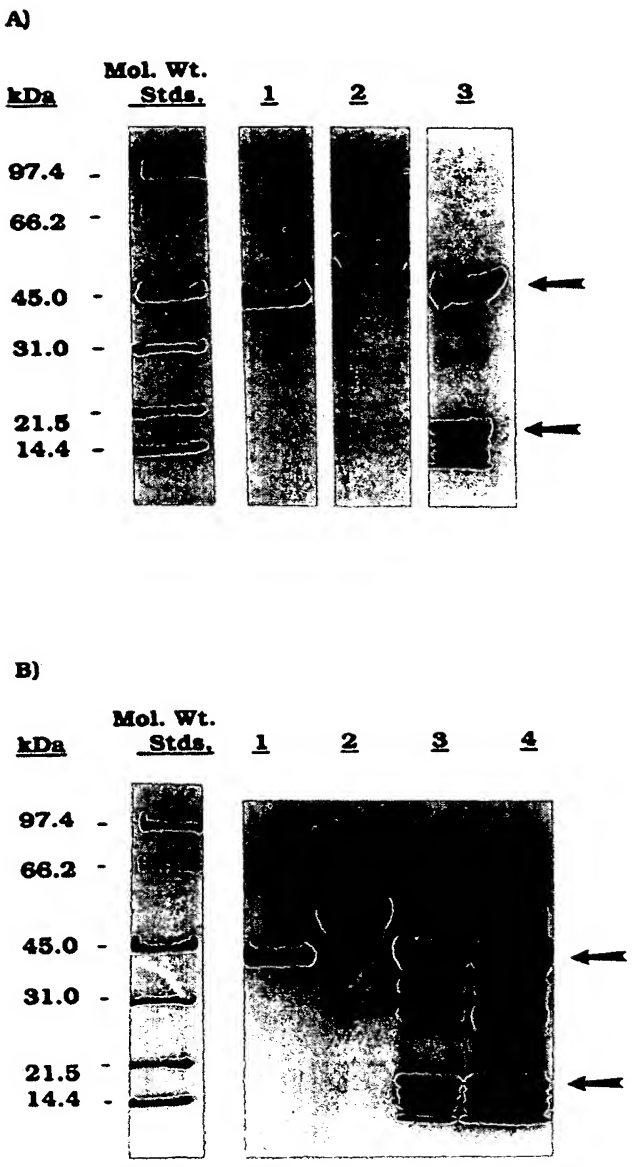


FIG. 9. **Factor Xa cleavage of the wild type MBP-NBF1 fusion protein.** *A*, the wild type fusion protein (MBP-NBF1) was treated with Factor Xa exactly as described under "Methods" and then subjected to SDS-PAGE. In all cases, 5 μg of protein was loaded onto the gel. *Lane 1*, untreated MBP control; *lane 2*, untreated MBP-NBF1 fusion protein control; *lane 3*, MBP-NBF1 treated with Factor Xa. *B*, separation of NBF1 from the Factor Xa cleavage mixture was attempted by binding the MBP component to an amylose resin. To do this the product from the cleavage reaction was mixed with the resin, incubated at room temperature for 10 min, and then centrifuged at 4000 × *g* for 10 min. The resultant pellet was washed with 20 mM HEPES, pH 7.5, and centrifuged again for 10 min at 4000 × *g*. Then, the proteins bound to the resin were eluted with 20 mM maltose containing 20 mM HEPES, 100 mM NaCl, and 2 mM $CaCl_2$, pH 7.5. Both the bound and unbound fractions (5 μg) were then analyzed by SDS-PAGE. *Lane 1*, untreated MBP control; *lane 2*, untreated MBP-NBF1 fusion protein control; *lane 3*; the amylose unbound fraction; *lane 4*, the amylose bound fraction. In both *A* and *B*, the *upper arrow* indicates the position of the MBP and the lower arrow the position of the cleaved NBF1. The identity of the NBF1 was confirmed by N-terminal sequencing of the band. Molecular size markers are the same as those indicated in the legend to Fig. 2. (See text for explanation of multiple bands in *A*, *lane 3*, and *B*, *lanes 3* and *4*.)

interactions among NBF1 domains of separate MDR or CFTR molecules. Removal of NBF1 domains alone from such an environment may allow their normal propensity to interact to go "unchecked" with polymer formation resulting. Second, the overexpressed NBF1 domain of CFTR (50, 51) and the nucleotide domain of the related bacterial HisP protein (52) have been reported to interact with the lipid bilayer of the membrane. This amphipathy of the domain must be compensated when placed in an aqueous buffer. Thus, polymer formation and, hence, stability could be promoted by interaction of the nonpolar regions of individual NBF1 molecules. Third, regardless of why such organized macromolecular assemblies of MBP-NBF1 fusion proteins form, the fact that they do form in aqueous buffer may prove useful in attempts to form crystals for structural analysis. Certainly, microcrystals containing TNP-ATP are readily formed in the presence of lithium sulfate and polyethylene glycol (Fig. 8), and efforts are currently under way in an attempt to grow large crystals suitable for x-ray diffraction analysis.

Finally, it should be noted that the studies reported here may have direct relevance to a new type of gene therapy that might be attempted in future studies to ameliorate the severity of cystic fibrosis caused by ΔF508 mutations. Thus, the introduction into airway epithelial cells of a cDNA encoding a small stable protein designed to target and interact specifically with the ΔF508,NBF1 domain may stabilize defective CFTR molecules restoring their normal function.

Acknowledgments—We are grateful to Dr. Wu-Schyong Liu for amino acid and N-terminal sequence analysis, Dr. L. Mario Amzel and to Dr. Mario Bianchet for help in crystallizing the fusion proteins, to Jody Franklin for preparation of oligonucleotides, and to Jackie Seidl for processing this manuscript for publication. We are grateful also for critical discussions with Drs. Michael Baumann, Krishan Arora, Saroj Mathupala, and Harry Price, as well as with Michael Lebowitz.

REFERENCES

1. Rommens, J. M. Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L., and Collins, F. S. (1989) *Science* **245**, 1059–1065
2. Riordan, J. R., Rommens, J. M., Karem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsik, J., Chou, J., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., and Tsui, L. (1989) *Science* **245**, 1066–1073
3. Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K. Chakravarti, A., Buchwald, M., and Tsui, L. (1989) *Science* **245**, 1073–1080
4. Walker, J. E., Saraste, M., Runswick, J., and Gay, N. J. (1982) *EMBO J.* **1**, 945–951
5. Cheng, S. H., Rich, D. P., Marshall, J., Gregory, R. J., Welsh, M. J., and Smith, A. E. (1991) *Cell* **66**, 1027–1036
6. Picciotto, M. R., Cohn, J., Bertuzzi, G., Greengard, P., and Nairn, A. C. (1992) *J. Biol. Chem.* **267**, 12742–12752
7. Cutting, G. R., Kasch, L. M., Rosenstein, B. J., Zielenski, J., Tsui, L., Antonarakis, S. E., and Karazian, H. H., Jr. (1990) *Nature* **346**, 366–369
8. Kerem, B., Zielenski, J., Markiewicz, D., Bozon, D., Gazit, E., Yahav. J., Kennedy, D., Riordan, J. R., Collins, F. S., Rommens, J. M., and Tsui, L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8447–8451
9. Cheadle, J. P., Meredith, A. L., and Aljader, N. (1992) *Hum. Mol. Genet.* **1**, 123–125
10. Anderson, M. P., Gregory, R. J., Thompson, S., Souza, A. E., Sucharita, P., Mulligan, R. C., Smith, A. E., and Welsh, M. J. (1991) *Science* **253**, 202–205
11. Rich, D. P., Gregory, R. J., Anderson, M. P., Manavalan, P., Smith, A. E., and Welsh, M. J. (1991) *Science* **253**, 205–207
12. McDonald, T. V., Nghiem, P. T., Gardner, P., and Martens, C. L. (1992) *J. Biol. Chem.* **267**, 3242–3248
13. Bear, C. E., Li., C., Kartner, N., Bridges, R. J., Jensen, T. J., Ramjeesingh, M., and Riordan, J. R. (1992) *Cell* **68**, 809–818
14. Tilly, B. C., Winter, M. C., Ostedgaard, L. S., O'Riordan, C., Smith, A. E., and Welsh, M. J. (1992) *J. Biol. Chem.* **267**, 9470–9473
15. Egan, M., Flotte, T., Afoine, A., Solow, R., Zeitlin, P. L., Carter, B. J., and Guggino, W. B. (1992) *Nature* **358**, 581–584
16. Drumm, M. L., Wilkinson, D. J., Smit, L. S., Worrell, R. T., Strong, T. V., Frizzell, R. A., Dawson, D. C., and Collins, F. S. (1991) *Science* **254**, 1797–1799
17. Anderson, M. P., Berger, H. A., Rich, D. P., Gregory, R. J., Smith, A. E., and Welsh, M. J. (1991) *Cell* **67**, 775–784
18. Quinton, P. M., and Reddy, M. M. (1992) *Nature* **360**, 79–81

19. Thomas, P. J., Shenbagamurthi, P., Ysern, X., and Pedersen P. L. (1991) *Science* **251**, 555–557
20. Thomas, P. J., Shenbagamurthi, P., Sondek, J., Hullihen, J. M., and Pedersen, P. L. (1992) *J. Biol. Chem.* **267**, 5727–5730
21. Thomas, P. J. Ko, Y. H., and Pedersen, P. L. (1992) *FEBS Lett.* **312**, 7–9
22. Denning, G. M., Anderson, M. P., Amara, J. F., Marshall, J., Smith, A. E., and Welsh, M. J. (1992) *Nature* **358**, 761–764
23. Li, C., Ramjeesingh, M., Reyes, E., Jensen, T., Chang, X., Rommens, J. M., Bear, C. E. (1993) *Nature Genet.* **3**, 311–316
24. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., and Holmes, K. C. (1990) *Nature* **347**, 37–44
25. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) *J. Biol. Chem.* **266** 5202–5219
26. MacFerrin, K. D., Terranova, M. P., Schreiber, S. L., and Verdine, G. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1937–1941
27. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., Arnheim, N. (1985) *Science* **230** 1350–1354
28. Guan, C., Li. P., Riggs, P. D., and Inouye, H. (1987) *Gene (Amst.)* **67**, 21–30
29. Maina, C. V., Riggs, P. D., Grandea, A. G., III, Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., and Guan, C. (1988) *Gene (Amst.)* **74**, 365–373
30. Johnston, T. C., Thompson, R. B., and Baldwin, T. O. (1986) *J. Biol. Chem.* **261**, 4805–4811
31. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene (Amst.)* **33**, 103–119
32. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
33. Tabor, S., and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4767–4771
34. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
35. Marston, F. A. O. (1987) in *DNA Cloning* (Glover, D., ed) Vol. 3, pp. 59–88, IRL Press, Oxford
36. Withold, B., Van Heerikhuizen, H., and De Leij, L. (1976) *Biochim. Biophys. Acta* **443** 534–544
37. Laemmli, U. K. (1970) *Nature* **227**, 680–685
38. Edman, P. (1950) *Acta. Chem. Scand.* **4**, 283–293
39. Hunkapiller, M. W., and Hood, L. E. (1983) *Science* **219**, 650–659
40. Hiratsuka, T., and Uchida, K. (1973) *Biochim. Biophys. Acta* **320**, 635–647
41. Hiratsuka, T. (1976) *Biochim. Biophys. Acta* **453**, 293–297
42. Hiratsuka, T. (1982) *Biochim. Biophys. Acta* **719**, 509–517
43. Mcpherson, A. (1990) *Eur. J. Biochem.* **189**, 1–23
44. Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254
45. Mullen, G. P., Shenbagamurthi, and Mildvan, A. S. (1989) *J. Biol. Chem.* **264**, 19637–19647
46. Chang, C. T., Wu, C.-S. C., Yang, J. T. (1978) *Anal. Biochem.* **91**, 13–31
47. Thomas, P. J., and Pedersen, P. L. (1993) *J. Bioenerg. Biomembr.* **25**, 11–19
48. Boscoboinik, D., Debanne, M. T., Stafford, A. R., Jung, C. Y., Gupta, R. S., and Epand, R. M. (1990) *Biochim. Biophys. Acta* **1027**, 225–228
49. Georges, E., Tsuruo, T., and Ling, V. (1993) *J. Biol. Chem.* **268**, 1792–1798
50. Hartman, J., Huang, Z., Rado, T. A., Peng, S., Jilling, T., Muccio, D. D., and Sorscher, E. J. (1992) *J. Biol. Chem.* **267**, 6455–6458
51. Arispe, N., Rojas, E., Hartman, J., Sorscher, E. J., and Pollard, H. P. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1539–1543
52. Mimura, C. S., Holbrook, S. R., and Ferro-Luzzi Ames, G. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 84–89

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 23, Issue of June 9, pp. 14085–14093, 1995
Printed in U.S.A.

Cloning, Overexpression, Purification, and Characterization of the Carboxyl-terminal Nucleotide Binding Domain of P-glycoprotein*

(Received for publication, August 26, 1994, and in revised form, January 24, 1995)

Sadhana Sharma‡ and David R. Rose

From the Division of Molecular and Structural Biology, Ontario Cancer Institute, and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M4X 1K9 Canada

Multidrug-resistant tumor cells overexpress P-glycoprotein (170 kDa), a member of the ABC (*ATP Binding Cassette*)-transporter superfamily. P-glycoprotein has been implicated in transport of a broad range of amphiphilic, hydrophobic drugs from tumor cells. The sequence and structural organization of P-glycoprotein, which consists of 12 transmembrane helices and two cytoplasmic nucleotide binding domains, is similar to other ABC-transporters. It is believed that the nucleotide binding domains of various ABC transporters, which have 30–50% sequence identity, play an important role in coupling ATP hydrolysis to the transport process. To allow structure-function studies of the nucleotide binding domains, the carboxyl-terminal nucleotide binding domain (NBD) of Chinese hamster P-glycoprotein has been cloned, overexpressed, and purified both by itself and as a fusion with maltose-binding protein. It has been demonstrated that the carboxyl-terminal NBD, when overexpressed in *Escherichia coli* in the absence of transmembrane helices, has very low ATPase activity. This suggests that the amino-terminal nucleotide binding domain and possibly interaction with the transmembrane domains may be required for full ATPase activity. It is also consistent with the idea that the ATPase activity of P-glycoprotein is stimulated in the presence of drugs. Circular dichroism spectral analysis and the ability of carboxyl-terminal NBD, both by itself and as a fusion with maltose-binding protein, to bind ATP-agarose beads and P-glycoprotein specific monoclonal antibodies suggests that the polypeptide folds into a functional domain. Gel filtration chromatography and cross-linking studies indicate that the carboxyl-terminal NBD has a tendency to self-associate to form oligomers. It is speculated that the carboxyl-terminal NBD may play a role in self-association of P-glycoprotein molecules in the plasma membrane.

Multidrug resistance, in which cancer cells manifest resistance not only to chemotherapeutic drugs but also to other structurally unrelated drugs, is a major clinical problem in cancer treatment (for reviews, see Refs. 1 and 2). This phenomenon is associated with the overexpression of the 170-kDa membrane-associated P-glycoprotein (Pgp).[1] This protein consists of two halves each containing six predicted hydrophobic membrane spanning segments followed by a cytoplasmic domain (3). The sequence and domain organization of Pgp is typical of the ABC (ATP binding cassette) superfamily of active transporters and is presumably the result of an early gene duplication (4). The two cytoplasmic domains of Pgp, which include the familiar Walker A and B motifs for nucleotide binding (5), show extensive sequence identity with the NBDs of other ABC-transporters (6, 7). Pgp functions as an energy-dependent drug efflux pump and thus plays a major role in the reduced accumulation of drugs inside the cell (8). Plasma membrane vesicles prepared from multidrug-resistant cells and reconstitution of purified Pgp in vesicles confirms that the drug transport is dependent on the constant supply of energy either from ATP or GTP (9–12). Pgp has also been associated with chloride (13) and ATP channel activities (14), and has peptide transport ability (1). ATP has been implicated in all the functions of Pgp.

Several residues have been mutated in the Walker A motif of both NBDs of Pgp with varied effects. Some mutations resulted in loss of transport activity but had no effect on binding to ATP analogues, whereas other mutations affected both transport and ATP binding (15). Also, a double mutation which alters both NBDs simultaneously shows reduced ATP channel activity (14). However, the mutations which alter the ATP hydrolyzing properties of Pgp have no effect on chloride channel activity. This suggests that unlike drug transport, the chloride channel function of Pgp does not require energy from ATP hydrolysis, although ATP binding is required for channel activation (13).

Several monoclonal antibodies recognizing either the extracellular or cytoplasmic domains of Pgp have been selected (1). Monoclonal antibody C219 is specific for a continuous short sequence present in both NBDs. Another monoclonal antibody, C494, recognizes only the carboxyl-terminal NBD. The binding of either C219 or C494 to Pgp inhibits the binding of ATP (16). Another antibody, JC66, which is specific for the distal carboxyl-terminal half of Pgp, completely inhibits the ATP channel activity of Pgp (14). The fact that C219 and C494 are equally effective in inhibiting ATP binding to Pgp and the inhibition of ATP channel activity by JC66 suggests that there is cooperativity between the NBDs.

The role of the NBDs in Pgp function, the structural changes upon ATP binding, and the coupling of ATP hydrolysis to the transport process remain poorly understood. This is primarily due to the complex structural organization and low abundance of the ABC-transporters (17, 18). Towards detailed structure-function study of the NBDs, we have cloned, expressed, purified, and characterized the carboxyl-terminal NBD of Chinese

* This work was supported by the National Cancer Institute of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 416-924-0671 (ext: 5053); Fax: 416-926-6529; E-mail: ssharma@oci.utoronto.ca.

[1] The abbreviations used are: Pgp, P-glycoprotein; ABC-transporters, *ATP binding cassette* transporters; NBD, nucleotide binding domain; TM, transmembrane; TMD, transmembrane domain; MBP, maltose-binding protein; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; DSP, dithiobis(succinimidyl propionate); DTSSP, 3,3′-dithiobis(sulfosuccinimidyl propionate); PCR, polymerase chain reaction; bp, base pair(s); IPTG, isopropyl-1-thio-β-D-galactopyranoside; PAGE, polyacrylamide gel electrophoresis; CFTR, cystic fibrosis transmembrane conductance regulator; FPLC, fast protein liquid chromatography.

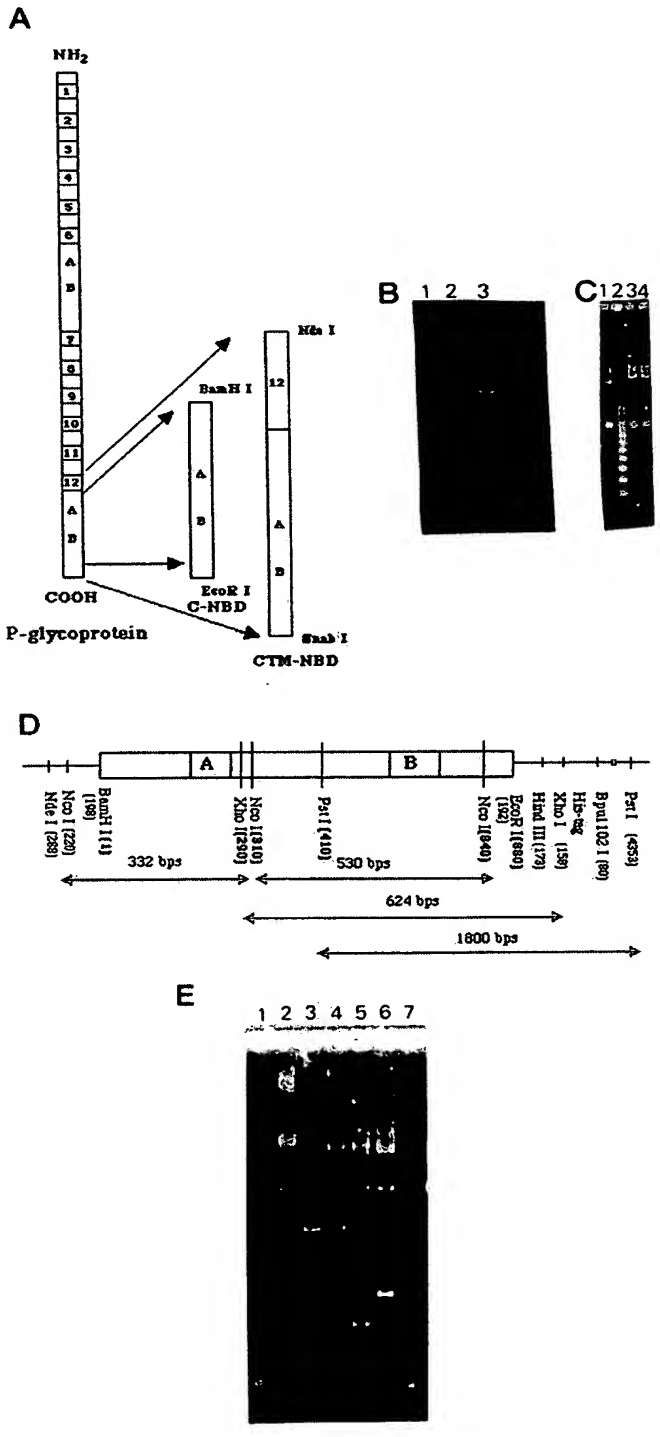


FIG. 1. **Amplification of the carboxyl-terminal NBD.** *A*, the Pgp is schematically represented with the transmembrane domains *boxed* and numbered 1–12 and two NBDs represented by *A* and *B* to denote Walker sequences. Two clones CTM-NBD and C-NBD were amplified. *B*, an agarose gel (1.0%) showing the amplified fragments. *Lane 1*, 100 base pair ladder; *Lane 2*, CTM-NBD; *Lane 3*, C-NBD. *C*, cloning of CTM-NBD in pT7–7. Various clones were picked, plasmid DNA was isolated. DNA was digested with restriction enzymes *Nde*I and *Bam*HI to confirm the presence of the insert. *Lanes 1, 3,* and *4* represent various clones screened for the CTM-NBD insert. *D*, restriction map of C-NBD. *E*, restriction digest analysis of C-NBD. *Lane 1*, markers (1057, 770, 612, 564, 495, 393, 345, and 341 bp); *Lane 2*, *Bam*HI; *Lane 3*, *Pst*I (1800 bp fragment); *Lane 4*, *Nco*I (530- and 330-bp fragments); *Lane 5*, *Xho*I (624 bp fragment); *Lane 6*, *Bam*HI-*Hin*dIII (880-bp fragment); *Lane 7*, same as *lane 1*.

hamster Pgp. Three different constructs were made in three different expression systems in an effort to optimize the yield and solubility of the purified protein. Expression as a fusion with maltose-binding protein (C-MBP) resulted in good protein yield, with the advantage that the known MBP structure (19) can aid in crystallographic structure determination of the carboxyl-terminal NBD of Pgp. Circular dichroism, and the ability of the expressed domains to bind ATP and antibodies, suggested that they retain a native-like fold. The fusion protein, C-MBP, retains relatively weak ATPase activity in the absence of transmembrane domains and the other NBD. The tendency of the carboxyl-terminal NBD to form oligomers suggests it may play an important role in self-association of Pgp in the plasma membrane.

EXPERIMENTAL PROCEDURES

Materials

A cDNA clone for Chinese hamster ovary Pgp was obtained from Dr. Victor Ling, Ontario Cancer Institute. Primers used for the polymerase chain reaction (PCR) were synthesized in the Ontario Cancer Institute Biotechnology Laboratory using an Applied Biosystems 392 DNA/RNA synthesizer. The pMal-c2 expression vector, amylose affinity resin, restriction enzymes, Vent DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs. The pT7–7 expression vector was kindly given to us by Dr. Bruce Waygood, University of Saskatchewan. The pET-22b(+) expression vector was obtained from Novagen. Ni-NTA-agarose and DNA purification kit were obtained from Qiagen Inc. ATP-agarose beads were purchased from Sigma. The DNA sequencing kit (Sequenase version 2.0 DNA) was from U. S. Biochemical Corp. A Pharmacia FPLC system was used for gel filtration.

Methods

Construction of the Expression Vector Containing Pgp Carboxyl-terminal NBD—Two different clones of the carboxyl-terminal NBD were amplified by PCR (Fig. 1A). CTM-NBD, which includes the last transmembrane (TM) helix, TM12 (M963) to the end of the transcript was cloned into the expression vector pT7–7 under the control of phage T7 RNA polymerase promoter (Fig. 2). The restriction enzyme sites, *Nde*I in the forward and *Sna*BI in the reverse PCR primer shown below, were incorporated to facilitate cloning into vector pT7–7 (20). Another construct, C-NBD, including amino acids Gln-984 to Arg-1276 but lacking TM12, was also amplified by PCR using the primers shown below with *Bam*HI and *Eco*RI restriction enzyme sites in the forward and reverse primer, respectively. C-NBD was cloned into pET-22b(+), also under the control of strong phage T7 transcription and translation signals (21). This expression vector contains a *pel*B leader sequence for the export of the target protein into the periplasm for improved folding and easier purification (22). Also at the COOH terminus of the protein there are six consecutive histidine residues. This hexahistidine tag facilitates purification through the high affinity of histidine for Ni-NTA resin (Qiagen Inc.), PCR-*Nde*I: 5′-GGGAACATATGACATTTGAAAATGTTC-TATTAGTAT-3′; PCR-*Sna*BI: 5′-GGGGTACGTACTTTCTTTTTAACT-TTGGTTAAATGC-3′; PCR-*Bam*HI: 5′-GGGGGGGATCCGCAGGTCA-GTTCATTTGCT-3′; PCR-*Eco*RI: 5′-GGGGGGAATTCGCGCTTTGCTC-CAGCCTGC-3′.

The polymerase chain reaction was carried out in a 100-μl reaction volume containing 20 mM Tris-HCl, pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 2.5 units of Vent DNA polymerase, 100 pmol of each primer, and 50 ng of template DNA which had been purified using QIAprep-spin plasmid purifications (Qiagen Inc.). The PCR reaction was carried out for 30 cycles and each cycle was performed as follows: denaturation at 92 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 30 s. The amplified PCR products, CTM-NBD (1000 bp) and C-NBD (880 bp), were purified using the QIAquick-spin PCR purification kit (Qiagen Inc.) and subsequently digested with appropriate restriction enzymes. The digested product was then subjected to agarose gel electrophoresis and the DNA was extracted using the QIAEX DNA gel extraction protocol (Qiagen Inc.).

The purified PCR product for CTM-NBD was ligated into the *Nde*I/*Sma*I restriction sites of expression plasmid pT7–7 and C-NBD was ligated into the *Bam*HI/*Eco*RI sites of plasmid pET-22b(+). Competent

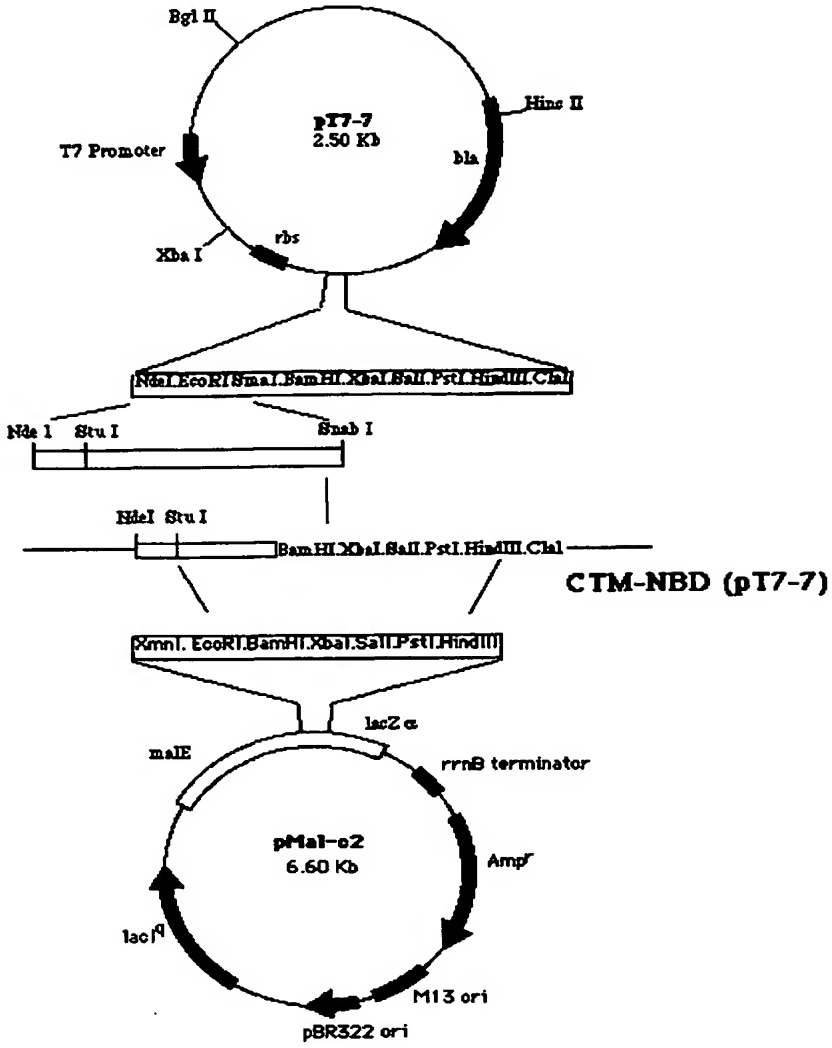


FIG. 2. **The schematic representation of cloning of carboxyl-terminal NBD in the expression vector.** The cloning into the expression vector pT7–7, pET22b(+), and subcloning into pMAL-c2 is diagrammatically represented. The amplified DNA fragment of CTM-NBD is ligated into *NdeI/Sma*I site of pT7–7 expression vector under the control of T7 promoter. The *StuI/Hin*dIII fragment from pT7–7 vector containing CTM-NBD fragment was ligated into *Eco*RI site which was blunt ended with Klenow fragment of DNA polymerase and the *Hin*dIII site in the multiple cloning site of pMal-c2 under the control of *tac* promoter. The insertion of the DNA fragment results in the disruptions of *lacZα* gene. When the plasmid containing the desired DNA fragment is transformed in lacα, complementing host white colonies are formed instead of blue in the presence of 5-bromo-4-chloro-3-indoyl β-D-galactoside and IPTG.

BL21(DE3)pLysS cells were transformed with the ligation mixture and clones were screened using PCR (23). The procedure is briefly described as follows: a single colony that is 1 mm in diameter was picked with a sterile pipette tip. Before transferring the colony to an Eppendorf tube containing 50 μl of sterile water, the pipette was touched to a YT+ampicillin (100 μg/ml) plate for later use. The tubes were placed in boiling water for 5 min and centrifuged for 1 min. A 10-μl aliquot of the supernatant was used for the PCR reaction using T7-promoter and terminator primer as described above.

To express the NBD as a fusion with maltose-binding protein, a *Stu*I-*Hin*dIII fragment (800 bp), including amino acid Leu-1023 to the end of the transcript, was subcloned from pT7–7 CTM-NBD into an *Eco*RI restriction enzyme site filled with the Klenow fragment of DNA polymerase and the *Hin*dIII restriction site of expression plasmid pMal-c2 (24) (Fig. 2). The insertion of the gene into the vector inactivates the β-galactosidase α-fragment activity of the *malE-lacZα* fusion and results in a blue to white color change on 5-bromo-4-chloro-3-indoyl β-D-galactoside and isopropyl-1-thio-β-D-galactopyranoside (IPTG) plates when the construct is transformed into an α-complementing host such as JM101 or NM522. The ligation mixture was transformed into competent NM522 *Escherichia coli* cells and grown on YT-agar plates containing ampicillin, 5-bromo-4-chloro-3-indoyl β-D-galactoside, and IPTG. The white colonies were screened by restriction digest analysis and subsequently sequenced using the Sanger dideoxy method (25) to confirm the fidelity of the PCR step.

Overexpression and Purification of CTM-NBD, C-NBD, and C-MBP—BL21(DE3)pLysS cells containing the appropriate construct, either CTM-NBD or C-NBD, were grown in 2 × YT broth containing 100 μg/ml ampicillin at 30 °C until A_{600} was 0.7. The culture was then induced with 1 mM IPTG and harvested 3 h later at 5000 × *g* for 20 min. After washing with 0.85% (w/v) NaCl, cells were resuspended in lysis buffer consisting of 20 mM Tris-HCl, pH 8.0, 10% glycerol, 500 mM NaCl, 1 mM phenylmethylsulfonyl fluoride, 10 mM β-mercaptoethanol, 0.1% Nonidet P-40, 10 μg/ml pepstatin, 10 μg/ml leupeptin, and 1% (v/v) aprotinin. The cells were frozen and kept at −20 °C until required. The frozen cells were thawed in a warm water bath and sonicated on ice for 2 × 5 min at a power setting 10 and 10% duty cycle with a 5-min interval using a Branson Model 450 sonifier. The lysed cells were centrifuged at 20,000 × *g* in a 45 Ti rotor (Beckmann) for 20 min. For C-NBD, supernatant was loaded onto a Ni-NTA-agarose column equilibrated with lysis buffer without protease inhibitors. After extensive washing, the protein was eluted with lysis buffer plus 200 mM imidazole. The fractions containing protein were detected by SDS-PAGE (26), pooled, and concentrated.

For CTM-NBD, the crude extract was loaded onto a Q-Sepharose (Pharmacia) column equilibrated with lysis buffer and the protein was eluted with a 0–1 M NaCl gradient. The fractions containing protein were concentrated and loaded onto a S-200HR (Pharmacia) gel filtration column also equilibrated with lysis buffer. After SDS-PAGE analysis, the fractions containing the protein of interest were pooled and concentrated.

For C-MBP, cells were grown, induced, harvested, and lysed in a similar fashion to those for CTM-NBD or C-NBD. The crude extract was loaded onto an amylose-affinity column equilibrated with lysis buffer. The column was extensively washed and the protein was eluted with the same buffer containing 10 mM maltose (27). Fractions were analyzed by SDS-PAGE, pooled, and dialyzed against 10 mM Tris-HCl, pH

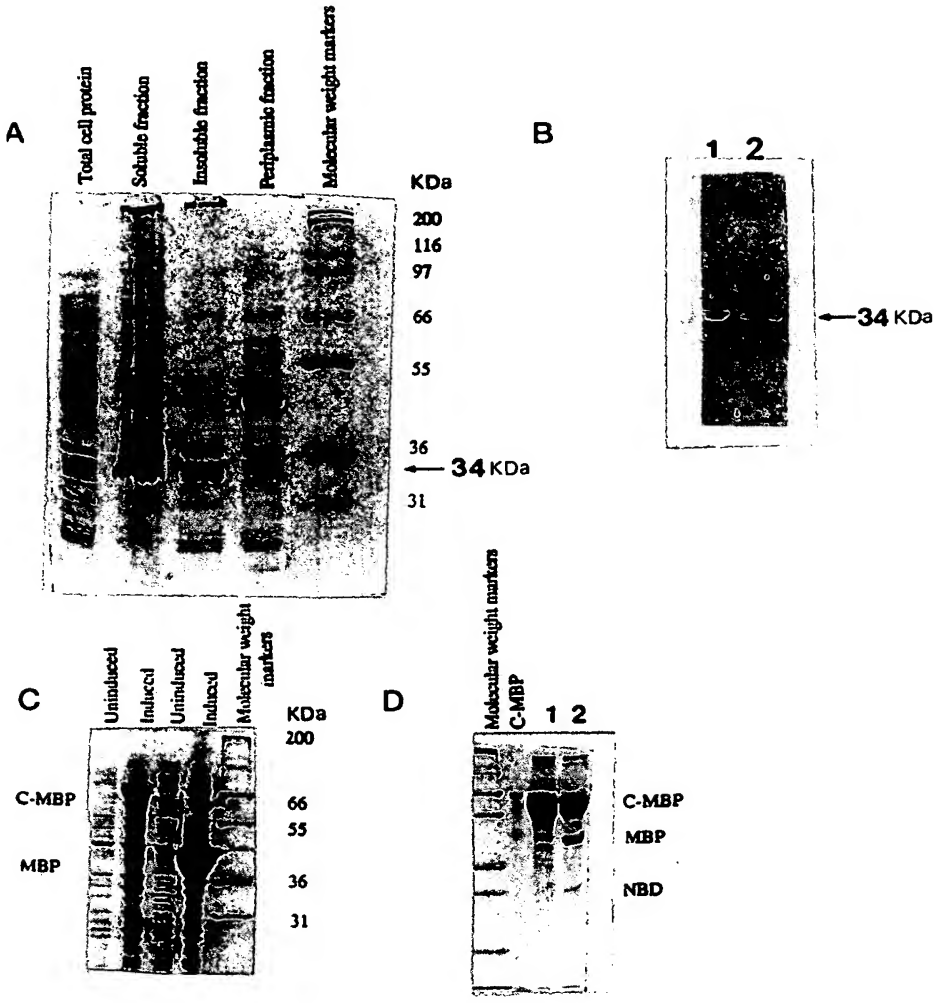


FIG. 3. **The SDS-PAGE gels for the analysis of overexpressed and purified protein.** BL21(DE3)pLysS cells containing either CTM-NBD or C-NBD were grown and induced as described under "Methods." *A*, the expression and the localization of CTM-NBD was checked in the various fractions (23). The expression of CTM-NBD appears to be low as shown in the total cell protein of induced cells and seems to be distributed equally in the soluble and insoluble fraction. There was no protein detected in the periplasmic fraction. *B*, the purified proteins, CTM-NBD (34 kDa) and C-NBD (34 kDa) are shown. *C*, the expression of C-MBP fusion protein. The total cell protein of uninduced and induced cells is compared. A protein band corresponding to ~68 kDa is induced in the presence of 1 mM IPTG, which correspond to C-MBP. The expression of MBP was used as a control and gel shows total cell protein before and after induction. *D*, the cleavage by specific protease Factor Xa. Factor Xa (200 μg/ml) was added to C-MBP (1 mg/ml) in 50 mM Tris-HCl buffer, pH 8.0, containing 500 mM NaCl and the reaction was carried at room temperature for various lengths of time. *Lanes 1* and *2* represent C-MBP before and after 24 h of digestion with Factor Xa.

TABLE I

Amino acid analysis was carried out as described under "Methods" on fusion protein C-MBP

The molar ratios are based on 98 mol of His/mol of fusion protein. The value for tryptophan is not reported because it gets degraded in presence of 6 M HCl and is not detectable. The low value of cysteine is also attributed to the degradation during hydrolysis.

Amino acid	Amount of amino acids	Theoretical value (from GCG)	Experimental value	Experimental/theoretical	Molar ratio
	nmol				
Asx[a]	7716.245	82	94	1.14	84.1
Glx[b]	6501.670	74	79	1.06	70.8
Ser	2286.068	30	28	0.9	24.9
Gly	5151.250	50	62	1.26	56.1
His	826.236	9	10	1.1	9.0
Arg	1876.950	19	22	1.2	20.4
Thr	2745.462	32	33	1.06	29.9
Ala	6270.499	64	76	1.2	68.3
Pro	3183.053	29	39	1.3	34.6
Tyr	1857.715	19	22	1.15	20.2
Val	3888.509	44	47	1.06	42.3
Met	879.015	9	10	1.1	9.6
Cys	60.900	3	1	0.3	0.7
Ile	3199.020	41	39	0.95	34.8
Leu	5209.918	57	63	1.1	56.7
Phe	1772.368	23	21	0.9	19.3
Lys	3394.617	51	42	0.82	36.9

[a] Asx; aspartic acid + asparagine.
[b] Glx; glutamic acid + glutamine.

8.0. The sample was then loaded onto a Q-Sepharose column equilibrated with the same buffer. Protein was eluted with a 0–1 M salt gradient in 10 mM Tris-HCl, pH 8.0. Appropriate fractions were pooled, concentrated, and the protein was exchanged into water using an Amicon ultrafiltration cell, with a YM10 membrane (Amicon Inc.).

Amino Acid Analysis of C-MBP—For amino acid analysis, 1000 pmol

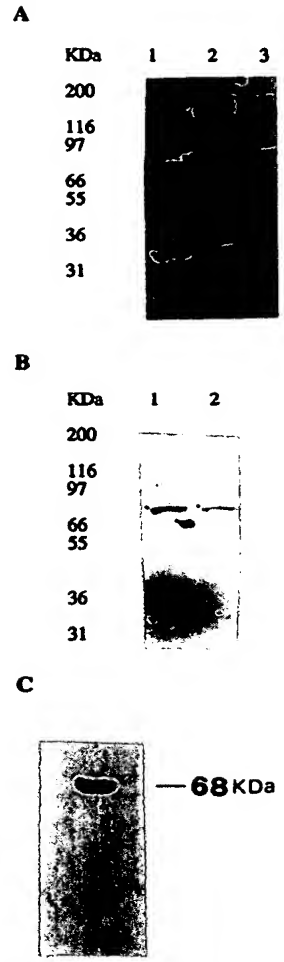


FIG. 4. **Detection of antibody binding by Western blot.** Western blot was performed as described under "Methods." The oligomeric forms corresponding to dimeric and trimeric forms of the carboxyl-terminal NBD can be detected by monoclonal antibody. *A*, detection of monoclonal antibody C219 binding to proteins expressed in *E. coli*. *Lane 1*, purified CTM-NBD (50 μg); *Lane 2*, purified C-NBD (50 μg); *Lane 3*, purified C-MBP (20 μg). *B*, binding of monoclonal antibody C494 to carboxyl-terminal NBD. *Lane 1*, purified CTM-NBD (20 μg); *Lane 2*, purified C-NBD (10 μg). *C*, detection of C-MBP (50 μg) with C494.

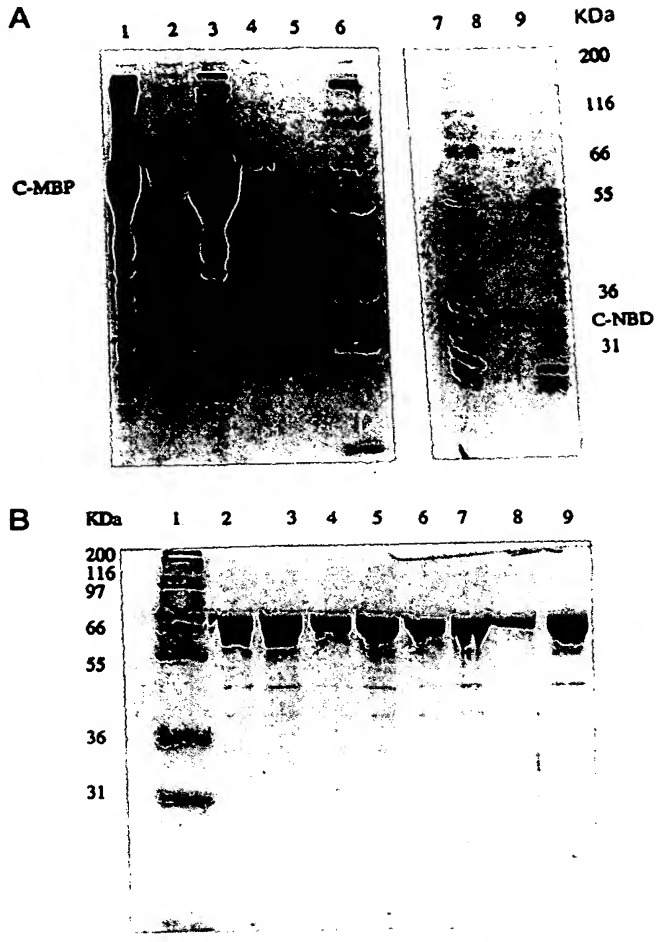


FIG. 5. **ATP binding characteristics of the expressed domains.** Partially purified proteins, C-MBP and C-NBD were incubated with agarose beads as described under "Methods." *A: Lane 1*, C-MBP (partially purified); *Lane 2*, C-MBP bound to the beads; *Lane 3*, wash; *Lane 4*, C-MBP (purified); *Lane 5*, C-MBP bound to the beads; *Lane 6*, molecular weight markers; *Lane 7*, C-NBD (partially purified); *Lane 8*, C-NBD bound to the beads; *Lane 9*, wash. *B*, binding competition with ATP. ATP (1–3 mM) was added to ATP-agarose beads prior to the incubation with C-MBP. *Lanes 2, 4, 6,* and *8* represent C-MBP bound to the beads in the presence of 0, 1, 2, and 3 mM ATP. *Lanes 3, 5, 7,* and *9* represent unbound protein.

of C-MBP was vapor-phase hydrolyzed in 6 N HCl for 48 h. The resulting amino acids were derivatized with phenyl isothiocynate and separated on a PICO-TAG (Pharmacia) column (3.8 mm × 15 cm) using reverse phase chromatography. The results were analyzed using PICO-TAG amino acid analysis software.

Western Blot Analysis—Western blotting was performed essentially according to Towbin *et al.* (28). Briefly, either crude extract or 20–100 μg of purified CTM-NBD, C-NBD, or C-MBP was mixed with an equal volume of gel loading buffer which lacked β-mercaptoethanol and electrophoresed on 10% SDS-PAGE. The protein was transferred to a polyvinylidene difluoride membrane using a Bio-Rad electroblotting apparatus. Polyvinylidene difluoride membranes were blocked for 2 h with 10% (w/v) non-fat dry milk in TBS buffer (0.1 M Tris-HCl, pH 7.2, 9% (w/v) NaCl) and then incubated for 1 h with 0.5 μg/ml monoclonal antibody C219 or C494 in the same buffer. The blot was thoroughly washed with TBS containing 1% (w/v) non-fat dry milk and 0.1% (v/v) Tween 20 and then incubated for 1 h with 1:2500 dilution of anti-mouse horseradish peroxidase-conjugated IgG in TBS buffer. After several washes with TBS buffer, the signal was detected by enhanced chemiluminescence (Amersham). The blot was exposed to Kodak X-Omat film for 15 s to 20 min.

ATP-affinity Column Binding—ATP-agarose beads (20 μl) were washed thoroughly with 10 mM Tris-HCl buffer, pH 8.0. Partially purified C-NBD or C-MBP was added to the beads and then incubated on a shaker for 1 h at 4 °C. Agarose beads were washed again with Tris-HCl buffer and then boiled in 30 μl of sample buffer to release protein bound to the beads. The samples were electrophoresed on 12% SDS-PAGE. ATP (1–3 mM) was used to compete the binding of protein to the agarose beads.

Circular Dichroism Spectroscopy (CD)—The fusion protein C-MBP (10 μM) was exchanged into water using an Amicon ultrafiltration cell with a YM30 membrane (Amicon Inc.). The CD spectrum for C-MBP was recorded between 200 and 250 nm at 23 °C in a 1-mm path cuvette using a Jasco J720 spectropolarimeter.

Gel Filtration FPLC—C-MBP (1 mg/ml) in 10 mM Tris-HCl, pH 8.0, was loaded onto a Pharmacia Superose 12 and Superose 6 FPLC gel filtration column equilibrated with the same buffer. The elution profiles of the C-MBP and standards, which consisted of blue dextran (200 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), and chymotrypsinogen A (25 kDa) were monitored at 280 nm. The elution volume was calculated and plotted *versus* $\log_{10}$ molecular weight. Gel filtration chromatography was also performed in the presence of 0.1% CHAPS in 10 mM Tris-HCl, pH 8.0.

Chemical Cross-linking—Chemical cross-linking of C-MBP was performed with dithiobis(succinimidyl propionate) (DSP) and 3,3′-dithiobis(sulfosuccinimidyl propionate) (DTSSP) (29). DSP was dissolved in *N,N*-dimethylformamide to give a final concentration of 2.5 mM and DTSSP was dissolved in 20 mM sodium phosphate buffer, pH 8.0, to a

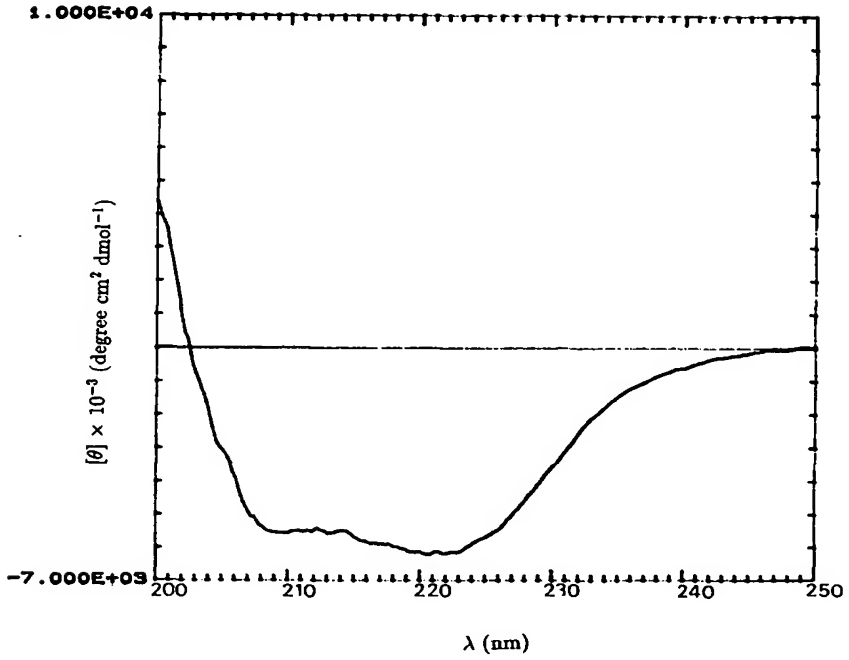


FIG. 6. **Circular dichroism spectra of C-MBP.** The fusion protein C-MBP (10 μM) and MBP (10 μM) was exchanged in water and the CD spectra was recorded as described under "Methods." The plot of molar ellipticity (θ) *versus* wavelength (nm) is depicted. *A*, spectra of MBP is shown with *dashed line* and C-MBP with *solid line*. *B*, difference spectrum which represents spectrum of NBD was obtained by subtracting spectra of C-MBP from MBP.

concentration of 0.116 mM. C-MBP (0.029 mM in 20 mM sodium phosphate, pH 8.0) was cross-linked with 0.0135–1.0 mM DSP and 0.0145–0.116 mM DTSSP in 20 mM sodium phosphate buffer. After 20 min at room temperature, the reactions were terminated by adding 50 mM glycine. The samples were loaded on 10% SDS-PAGE after the addition of an equal volume of gel-loading buffer which lacked β-mercaptoethanol.

Measurement of Mg^{2+}-ATPase Activity Using a Colorimetric Assay—Mg^{2+}-ATPase activity was determined by measuring the release of inorganic phosphate using a colorimetric assay (30). The assay solution contained 100 μg of C-MBP in 360 μl of assay buffer (50 mM Tris-HCl, pH 7.6, 0.15 mM NH_4Cl, 5 mM $MgCl_2$, 2 mM ouabain, 100 μM EGTA, 2 mM dithiothreitol, and 0.89 mM CHAPS). The reaction was initiated by adding 40 μl of ATP to give a final concentration of 2 mM in a 400-μl final volume. After 1 h at 37 °C, the reaction was terminated by adding 400 μl of 6% SDS, 3% ascorbate, and 0.5 M sodium molybdate in 0.5 M HCl. Products were stabilized by adding 400 μl of 2% sodium citrate, 2% sodium meta-arsenite, and 2% acetic acid. After incubation at 37 °C for 10 min the absorbance was read at 850 nm.

RESULTS

Overexpression and Purification of CTM-NBD, C-NBD, and Fusion Protein C-MBP—The amplified PCR products for CTM-NBD and C-NBD (Fig. 1*B*) were cloned into pT7–7 (Fig. 1*C*) and pET-22b(+) expression vectors under the control of a strong bacteriophage T7 promoter and the *Stu*I-*Hin*dIII fragment from CTM-NBD was subcloned in-frame with maltose-binding protein in the pMal-c2 expression vector (Fig. 2). The correct orientation of the clones was verified by sequencing and restriction digest analysis (Fig. 1, *D* and *E*). The IPTG-inducible expression of CTM-NBD and C-NBD and the amount of protein in the soluble fraction were moderate (Fig. 3*A*). Also the protein repeatedly precipitated during purification. The overall yield of purified protein was 25–50 μg/liter of cell culture. However, when expressed as a fusion with MBP, there was a significant increase in expression level and most of the protein remained soluble (Fig. 3*C*). Using amylose resin, for which MBP has great affinity, 80 mg of fusion protein was purified from a liter of cell culture. Purified CTM-NBD, C-NBD, and the fusion protein C-MBP were >90% pure as judged by SDS-PAGE (Fig. 3, *B* and *D*). The number of amino acids determined by amino acid analysis of C-MBP agreed closely with the theoretical value calculated using the program GCG (Genetics Computer Group, Wisconsin), thus confirming the identity and purity of the fusion protein (Table I).

Expression vector pMal-c2 contains the sequence coding for the recognition site of the specific protease Factor Xa, located 3′ of MBP, allowing cleavage of MBP from the protein of interest. Cleavage with Factor Xa for C-MBP was very slow: after 24 h only 20% of the protein was cleaved, suggesting that the protease site may have low accessibility (Fig. 3*D*).

Western Blot Analysis—The purified proteins were checked for their ability to bind to Pgp-specific monoclonal antibodies, C219 (Fig. 4*A*) and C494 (Fig. 4, *B* and *C*), by Western blot analysis. C219 and C494 recognize all three carboxyl-terminal NBD clones, *i.e.* CTM-NBD, C-NBD, and C-MBP, expressed in *E. coli*. MBP alone has no affinity for the antibodies (data not shown). Binding of C-MBP to C219 and C494 was also confirmed by enzyme-linked immunoassay. These results further confirm that the purified protein is the NBD of Pgp and not any other protein from *E. coli* of the same molecular mass.

ATP-agarose Binding Characteristics—CTM-NBD and C-MBP both bind ATP-agarose beads suggesting that the NBD retains its secondary and tertiary structure (Fig. 5*A*). In a control experiment, MBP alone did not bind the beads indicating that the interactions with ATP are due to the NBD alone. The binding properties of C-MBP, C-NBD, and CTM-NBD were the same. All showed decreased binding to ATP-agarose beads in the presence of ATP (Fig. 5*B*). CTM-NBD and C-MBP bind to an Affi-Gel® Blue affinity gel column (Bio-Rad) (results not shown), which is specific for nucleotide-binding proteins. This further indicates a correctly folded protein that retains ATP binding ability.

Secondary Structure of C-MBP as Determined by CD—The ability of the carboxyl-terminal NBD to bind to ATP suggests that it retains a functional tertiary structure. This was further confirmed by CD spectral analysis. A CD spectrum was recorded for C-MBP between 200 and 250 nm. The plot of mean residue molar ellipticity of C-MBP exhibits two minima at 209 and 220 nm (Fig. 6). From the CD spectrum, C-MBP was calculated to consist of 14% α-helix, 42% β-strand, 20% turn, and 23% irregular structure. This secondary structure calculation was carried out using Prosec software package. Thus, C-MBP is highly structured.

Oligomer Formation—The tendency of the carboxyl-terminal

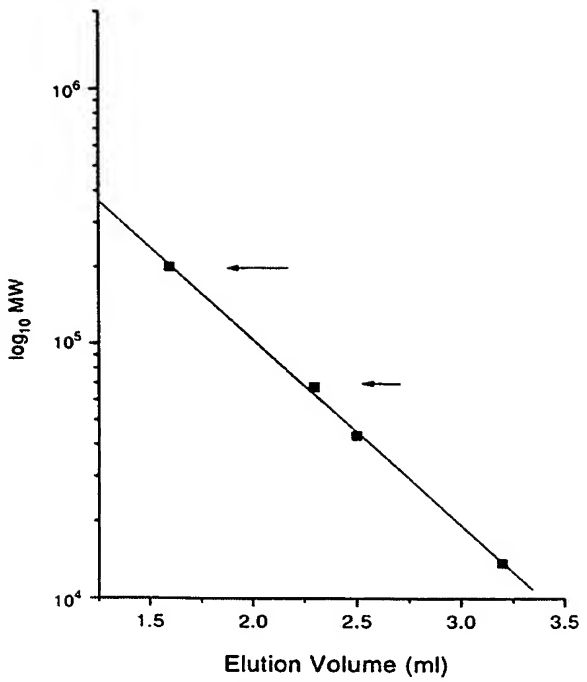


FIG. 7. **Molecular weight by gel filtration FPLC.** The elution volume of the standards which includes blue dextran, bovine serum albumin, ovalbumin, and chymotrypsinogen A was determined and plotted as a function of $\log_{10}$ molecular weight. Elution volume of purified C-MBP was determined under the same conditions. Molecular mass of C-MBP was estimated to be approximately 2000 and 68 kDa shown by *arrows* on the graph.

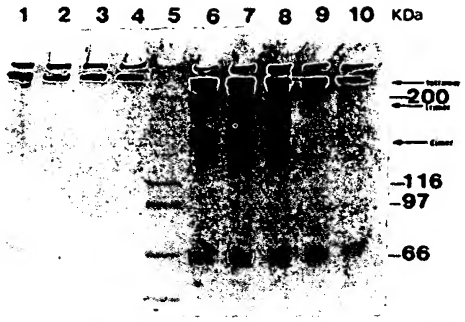


FIG. 8. **Identification of intermediate oligomeric forms of C-MBP by chemical cross-linking with DSP and DTSSP.** A series of cross-linking experiments were performed with purified 0.029 mM C-MBP as described under "Methods." *Lanes 1–4* shows cross-linking with increasing concentration of DSP (0.0135, 0.135, 0.5, and 1.0 mM); *lane 5* is molecular weight markers; *Lanes 7–10* represent cross-linking with increasing concentration of DTSSP (0.0073, 0.015, 0.029, 0.058, and 0.116 mM).

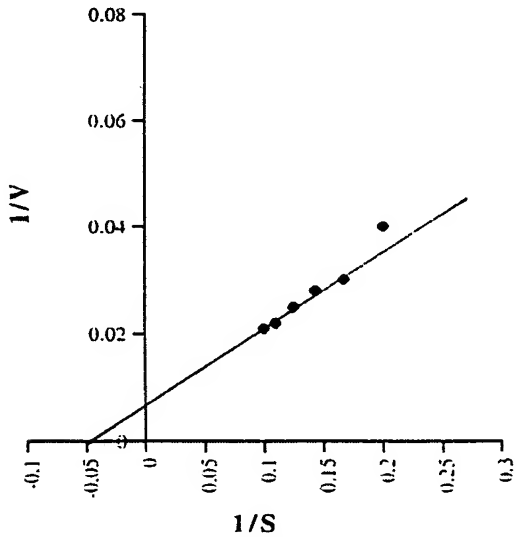


FIG. 9. **Lineweaver-Burk plot of dependence of ATPase activity of fusion protein C-MBP.** Assay buffer contained 100 μg of C-MBP in 400 μl of 400 μl of 50 mM Tris-HCl, pH 7.6, 0.15 mM NH_4Cl, 60 mM $MgCl_2$, 2 mM ouabain, 100 μM EGTA, 2 mM dithiothreitol, 0.89 mM CHAPS, and various concentrations of ATP (1–10 mM). V_{max} and K_m was calculated from the plot of 1/V (1/nmol of ATP hydrolyzed per min per mg of C-MBP) *versus* 1/S (1/[ATP] mM). V_{max} of 23.8 nmol/min/mg and a K_m of 20 mM for ATP was calculated from the plot.

NBD to form oligomers was evident from gel filtration and chemical cross-linking experiments. In the former, most C-MBP elutes at the void volume (with the blue dextran) of an FPLC gel filtration column, exhibiting a molecular mass of 2000 kDa. However, a small fraction of protein also elutes with bovine serum albumin which has about the same M_r as C-MBP (Fig. 7). Performing gel filtration chromatography in the presence of 0.1% (v/v) CHAPS had no effect on the elution profile of the standards and C-MBP (results not shown). In a control gel filtration experiment involving MBP alone, there was no evidence for oligomerization. The tendency of C-MBP to oligomerize was further evidenced by using the homobifunctional, thio-cleavable and amine-reactive chemical cross-linkers DSP and DTSSP. With such reagents, proteins in close proximity have the potential to be chemically cross-linked. With a low concentration of DTSSP (0.0073, 0.0145, and 0.029 mM), bands which correspond to dimeric (136 kDa) and trimeric (204 kDa) forms of C-MBP are seen (Fig. 8). With DSP, such intermediate forms could not be detected and most protein was present as higher order oligomers. MBP alone did not form oligomers with either DTSSP or DSP. Furthermore, Western blot analysis confirms that the Pgp-specific monoclonal antibodies are able to recognize the oligomeric forms of CTM-NBD, C-NBD, and C-MBP (Fig. 4).

ATPase Activity of the Fusion Protein—C-MBP exhibits very low Mg^{2+}-dependent ATPase activity with a V_{max} of 23.8 nmol/min/mg and a K_m of 20 mM as determined by a Lineweaver-Burk plot (Fig. 9). The presence of ouabain (2 mM) and EGTA (100 μM) has no effect on the activity, suggesting that the observed ATPase activity is not due to the presence of contaminating Na^+-K^+- or Ca^{2+}-ATPase. ATPase activity is dependent on the concentration of ATP and Mg^{2+} in the assay buffer. In the presence of 5 mM $MgCl_2$, ATPase activity decreased with increasing concentration of ATP. The observed decrease may be due to a limiting concentration of Mg^{2+} in the assay buffer, since no decrease in activity with increasing ATP concentration was observed when the assay was performed in the presence of 60 mM $MgCl_2$. In the reverse experiment, in which ATP concentration was kept constant, a similar dependence of ATPase activity on Mg^{2+} concentration was observed (Fig. 10): in the presence of 2 mM ATP, ATPase activity decreased as Mg^{2+} concentration increased. The observed decrease can be attributed to a limiting concentration of ATP. The kinetic parameters of the carboxyl-terminal NBD, in the absence of the Pgp transmembrane helices, are different from those of intact Pgp. The ATPase activity of Chinese hamster Pgp decreases in the presence of high concentrations of ATP with maximal activity at 6 mM ATP (12). In some cases a higher Mg^{2+} concentration (50 mM) caused inhibition of ATPase activity and in other cases there was no inhibition up to 50 mM Mg^{2+} (12).

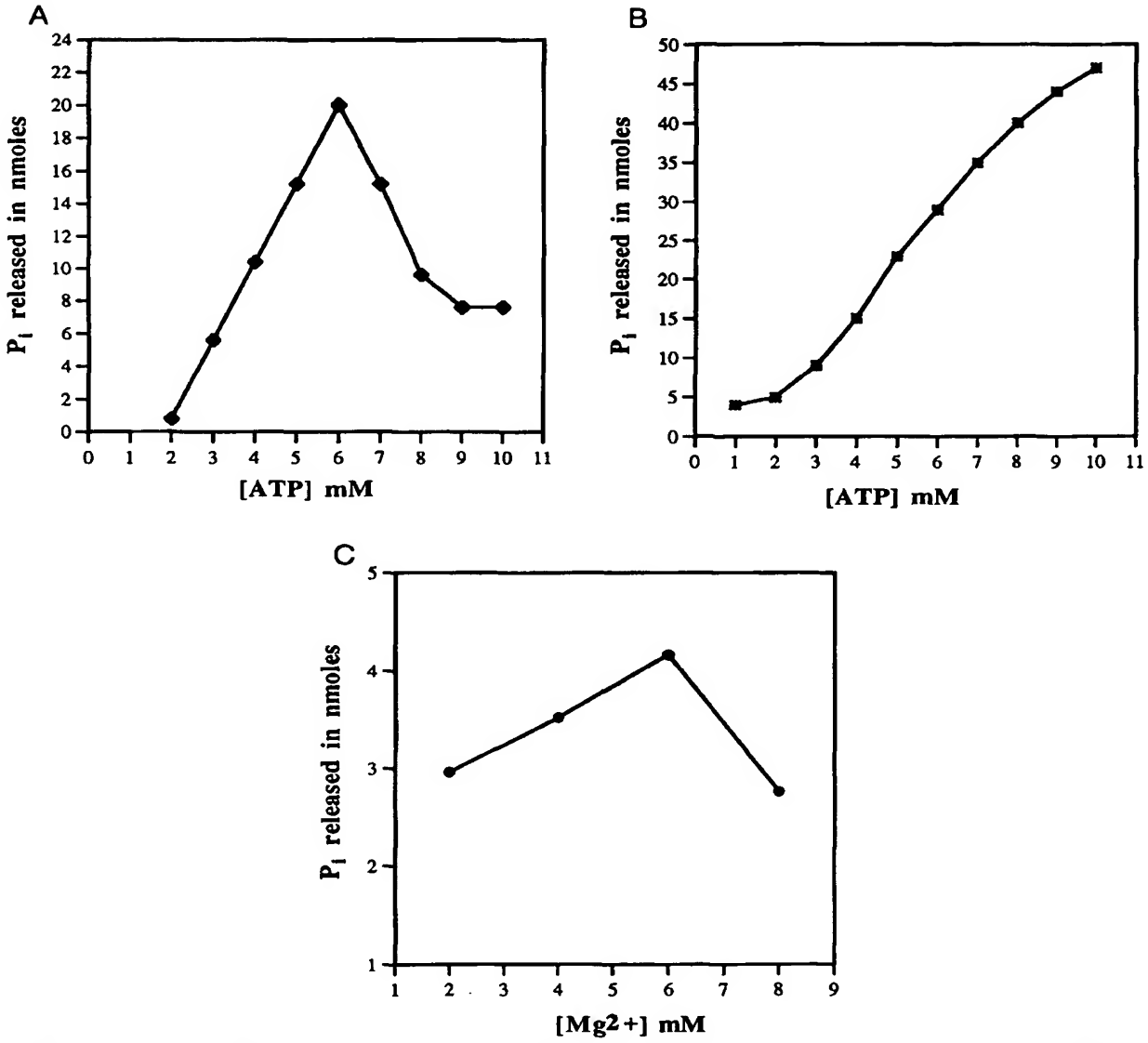


FIG. 10. **Dependence of ATPase activity on concentrations of $MgCl_2$ and ATP.** Each assay sample contained 100 μg of C-MBP in 400 μl of 50 mM Tris-HCl, pH 7.6, 0.15 mM NH_4Cl, 5 mM $MgCl_2$, 2 mM ouabain, 100 μM EGTA, 2 mM dithiothreitol, and 0.89 mM CHAPS or as indicated. *A*, ATP dependence in the presence of 5 mM $MgCl_2$. *B*, ATP dependence in the presence of 60 mM $MgCl_2$. *C*, Mg^{2+} dependence in the presence of 3 mM ATP.

DISCUSSION

A central question in multidrug resistance is how a single integral membrane protein, Pgp, can transport a wide variety of drugs and hydrophobic peptides. The role of ATP and how ATP-derived energy is harnessed in the transport process of Pgp and other ABC-transporters remains unknown. It has been suggested that the transport of solutes by an ABC-transporter across the membrane is dependent on energy derived from ATP and that two molecules of ATP are hydrolyzed per transport event (31). The latter proposal is consistent with the fact that most ABC-transporters contain two NBDs. The hemolysin and glycine-betaine transporters, which have only one NBD, are known to function as dimers (17). The NBDs from several ABC-transporters, for example, cystic fibrosis transmembrane conductance regulator (CFTR) and maltose (MalK), histidine (HisP), and oligopeptide (OppF) permeases, have been overexpressed and purified (32–36). The purified domains are able to bind ATP and its analogues but show no ATPase activity. It has also been suggested that these nucleotide-binding domains may not be able to hydrolyze ATP in the absence of the membrane binding components of the transporters (31).

In this study, we have chosen the carboxyl-terminal NBD of Pgp as a model system for structural and functional characterization of ABC-transporter NBDs. Early reports suggest that, like any other ABC-transporter, Pgp requires both of its NBDs for efficient activity and also that the ATPase activity is stimulated in the presence of drugs (33). Georges *et al.* (1991) suggested that there is cooperativity between the two NBDs of Pgp, as evidenced by the study of mutants having deletions in the amino or carboxyl half of Pgp. Deletion mutations in either domain resulted in the loss of transport function, indicating that both halves are necessary (37, 38). This contradicts the finding that the amino-terminal NBD alone contains all the residues necessary to hydrolyze ATP (39). These studies were carried out

with either intact Pgp or the isolated amino-terminal half including TMDs 1–6. To investigate the role of the carboxyl-terminal NBD of Pgp in transport, chloride, and ATP channel activities, we have overexpressed this domain in *E. coli*.

When overexpressed, the CTM-NBD and C-NBD proteins were found in the membrane fraction, indicating that either the protein is very hydrophobic or is associated with the membrane. This is not surprising for CTM-NBD since it contains the amino acid residues from TM12. C-NBD, which lacks TM12 residues, was expressed in pET22b(+). This expression plasmid contains a *pelB* leader sequence, which directs proteins to the periplasmic space. Since C-NBD also remained associated with the membrane, it might be speculated that in intact Pgp, the NBDs somehow associate with the membrane. Similar observations have been made with the NBDs of bacterial permeases such as HisP, OppF, and MalK and with those of human CFTR. When expressed in the absence of TMDs, these NBDs remain closely associated with the membrane (32–36). Although the NBDs and the TMDs of bacterial ABC-transporters are encoded by different sets of genes, the proteins tightly associate with each other to form the intact transporter. It is interesting to speculate how NBDs which are highly hydrophilic might interact with the membrane. It was proposed that a segment of NBD protrudes into or through a pore generated by the TMDs (35). However, this remains to be verified by, for example, three-dimensional structural information.

Expression of the carboxyl-terminal NBD as a fusion with MBP showed tremendous improvement in the protein yield and solubility. There are additional advantages to expressing protein as a fusion with MBP: the affinity of MBP for amylose can be used for rapid purification and the three-dimensional structure of MBP has been elucidated (19). This could yield information useful in the ultimate structure determination of the Pgp NBD. The inability to efficiently cleave MBP from the NBD with Factor Xa suggests that the protease site has low accessibility. A similar observation was made with the NBD of CFTR when expressed as a fusion with MBP. It was speculated that the CFTR NBD may interact strongly with MBP (33).

CTM-NBD and C-MBP retain the ability to bind ATP. Unlike NBDs of other ABC-transporters, the carboxyl-terminal NBD of Pgp retains low ATPase activity when expressed in the absence of the other NBD and TMDs. Previous work suggests that upon mutating one of the two NBDs, Pgp loses its activity, indicating that the two NBDs function in a cooperative manner (37). Interactions of the carboxyl-terminal NBD with either or both of the TMDs and the amino-terminal NBD of Pgp may be required for full ATPase activity. Inhibition of ATPase activity with a high concentration of both Mg^{2+} and ATP, as observed with intact Pgp, was not observed for the carboxyl-terminal NBD. This suggests that some regulation of the ATPase activity of intact Pgp is lost when the NBD is expressed in the absence of the TMDs and the other NBD. These results are also consistent with the fact that ATPase activity of intact Pgp is stimulated in the presence of drugs.

The NBDs of Pgp share considerable sequence similarity with the NBDs of CFTR. For example, there is 32% identity and 56% similarity between the carboxyl-terminal NBD of Pgp and CFTR nucleotide binding fold 1. However, the carboxyl-terminal domain of Pgp shows rather weak binding to TNP-ATP, a fluorescent analogue of ATP, unlike nucleotide binding fold 1 of CFTR (33). Such differences in ATP binding behavior of these domains possibly evidences subtle active-site structural differences.

It has previously been observed that Pgp exists as a dimer or trimer in the plasma membrane (40, 41). The results from this study, showing that CTM-NBD, C-NBD, and C-MBP tend to form dimers, trimers, and higher oligomers, suggest that the carboxyl-terminal NBD may play an important role in the self-association of Pgp within the plasma membrane. One can speculate from these results that intact Pgp molecules undergo ordered self-association through interactions between NBDs. However, an NBD in the absence of TMDs and the other NBD may undergo less ordered self-association with the resulting tendency to form higher oligomers.

The successful expression and purification of the carboxyl-terminal NBD of Pgp is a critical step in structure-function studies of the molecule. In combination with an analogous approach to the expression and characterization of the amino-terminal NBD, we hope to improve understanding of the role of the NBDs in the function of ABC-transporters.

Acknowledgments—Drs. S. Bagby, M. Ikura, and P. Yau are thanked for their help, encouragement, and scientific discussions. S. Bagby's help in preparing and critical reading of the manuscript is greatly appreciated. We acknowledge the help of Dr. A. Shapiro with ATPase assays and Dr. A. Chakrabartty with CD analysis.

REFERENCES

1. Endicott, J., and Ling, V. (1989) *Annu. Rev. Biochem.* **58**, 137–171
2. Gottesman, M. M., and Pastan, I. (1993) *Annu. Rev. Biochem.* **62**, 385–427
3. Chen, C. J., Chin, J. E., Ueda, K., Clark, D. P., Pastan, I., Gottesman, M. M., and Roninson, I. B. (1986) *Cell* **47**, 381–389
4. Gros, P., Croop, J., and Housman, D. E. (1986) *Cell* **47**, 371–390
5. Walker, J. E., Sarste, M., Runswick, M. J., and Gay, N. J. (1982) *EMBO J.* **1**, 945–951
6. Hyde, S. C., Emsley, P., Hartshon, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R., Gallagher, M. P., Gill, D. R., Hubbard, R. E., and Higgins, C. F. (1990) *Nature* **346**, 362–365
7. Mimura, C. S., Holbrook, S. R., and Ames, G. F. L. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 84–88
8. Kessel, D., Botterill, V., and Wodinsky. I. (1968) *Cancer Res.* **28**, 938–941
9. Lelong, I. H., Padmanabhan, R., Lovelace, E., Pastan, I., and Gottesman, M. M. (1992) *FEBS Lett.* **304**, 256–260
10. Al-Shawi, M., and Senior, A. E. (1993) *J. Biol. Chem.* **268**, 4197–4206
11. Sharom, F. J., Xiaohong, Y., and Doige, C. A. (1993) *J. Biol. Chem.* **268**, 24197–24202
12. Shapiro, A. B., and Ling, V. (1994) *J. Biol. Chem.* **269**, 3745–3754
13. Gill, D. R., Hyde, S. C., Higgins, C. F., Valverde, M. A., Mintenig, G. M., and Sepulveda, F. V. (1992) *Cell* **71**, 23–32
14. Abraham, E. H., Prat, A. G., Gerweck, L., Seneveratne, T., Arceci, R. K., Guidotti, G., and Cantiello, H. F. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 312–316
15. Azzaria, M., Schurr, E., and Gros, P. (1989) *Mol. Cell. Biol.* **9**, 5289–5297
16. Georges, E., Zhang, J.-T., and Ling, V. (1991) *J. Cell. Physiol.* **148**, 479–484
17. Higgins, C. F., Hyde, S. C., Mimmack, M. M., Gileadi, U., Gill, D. R., and Gallgher, M. P. (1990) *J. Bioenerg. Biomembr.* **22**, 571–592
18. Higgins, C. F. (1992) *Annu. Rev. Cell Biol.* **8**, 67–113
19. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) *J. Biol. Chem.* **266**, 5202–5219
20. Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1074–1078
21. Studier, F., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89
22. Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J., and Wilcox, G. (1987) *J. Bacteriol.* **169**, 4379–4383
23. Novagen, Inc. (1992) *pET System Manual*, Novagen Inc. (company newsletter)
24. Guan, C., Li, P., Riggs, P. D., and Inouye, H. (1987) *Gene (Amst.)* **67**, 21–30
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
26. Laemmli, U. K. (1970) *Nature* **227**, 680–685
27. Maina, C. V., Riggs, P. D., Grandea, A. G., III, Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., and Guan, C. (1988) *Gene (Amst.)* **74**, 365–373
28. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354
29. Maman, J. D., Yager, T. D., and Allan, J. (1994) *Biochemistry* **33**, 1300–1310
30. Clifflet, S., Torriglia, A., Chiesa, R., and Tolosa, S. (1988) *Anal. Biochem.* **168**, 1–4
31. Mimmack, M. L., Gallagher, M. P., Pearce, S. R., Hyde, S. C., Booth, I. R., and Higgins, C. F. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8257–8261
32. Hartman, J., Huang, Z., Rado, T. A., Peng, S., Jilling, T., Muccio, D. D., and Sorscher, E. J. (1992) *J. Biol. Chem.* **267**, 6455–6458
33. Ko, Y. H., Thomas, P. J., Delannoy, M. R., and Pederson, P. L. (1993) *J. Biol. Chem.* **268**, 24330–24338
34. Shuman, H. A., and Silhavy, T. J. (1980) *J. Biol. Chem.* **256**, 560–562
35. Kerppola, R. E., Shyamala, V. K., Klebba, P., and Ames, G. F. (1991) *J. Biol. Chem.* **266**, 9857–9865
36. Gallagher, M. P., Pearce, S. R., and Higgins, C. F. (1989) *Eur. J. Biochem.* **180**, 133–141
37. Azzaria, M., Schurr, E., and Gros, P. (1989) *Mol. Cell. Biol.* **9**, 5289–5297
38. Ambudkar, S. V., Lelong, I. H., Zhang, J., Cardarelli, C. O., Gottesman, M. M., and Pastan, I. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8472–8476
39. Shimabuku, A. M., Nishimoto, T., Ueda, K., and Komano, T. (1992) *J. Biol. Chem.* **267**, 4308–4311
40. Boscoboinik, D., Debanne, M. T., Stafford, A. R., Jung, C. Y., Gupta, R. S., and Epand, R. M. (1990) *Biochim. Biophys. Acta* **1027**, 225–228
41. Poruchynsky, M. S., and Ling, V. (1994) *Biochemistry* **33**, 4163–4174

Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: comparison of fusions to glutathione S-transferase, thioredoxin and maltose-binding protein

Changsen WANG, Ariel F. CASTRO, Denise M. WILKES and Guillermo A. ALTENBERG[1]

Department of Physiology and Biophysics, Basic Science Building, The University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0641, U.S.A.

Many membrane proteins that belong to the ATP-binding cassette (ABC) superfamily are clinically important, including the cystic fibrosis transmembrane conductance regulator, the sulphonylurea receptor and P-glycoprotein (multidrug resistance gene product; MDR1). These proteins contain two multispanning transmembrane domains, each followed by one nucleotide-binding domain (NBD) and a linker region distal to the first NBD. ATP hydrolysis by the NBDs is critical for ABC protein function; the linker region seems to have a regulatory role. Previous attempts to express soluble NBDs and/or linker regions without detergent solubilization, or to purify NBDs at high yields as soluble fusion proteins, have been unsuccessful. Here we present a system for the expression in *Escherichia coli* of the first NBD of MDR1 followed by its linker region (NBD1-MLD). A comparison of the expressions of NBD1MLD fused to glutathione S-transferase, thioredoxin and maltose-binding protein (MBP) shows that a high level of expression in the soluble fraction (approx. 8% of total *E. coli* protein) can be achieved only for MBP–NBD1MLD. The addition of a proteolytic thrombin site just proximal to the N-terminal end of NBD1MLD allows the cleavage of NBD1MLD from MBP, which can be easily purified with retention of its ATPase activity. In summary, success was obtained only when using an MBP fusion protein vector containing a thrombin proteolytic site between MBP and NBD1MLD. The approach described here could be generally applicable to solving the problems of expression and purification of NBDs/linker regions of ABC proteins.

Key words: cancer, cystic fibrosis, fusion proteins, multidrug resistance, P-glycoprotein.

INTRODUCTION

Membrane proteins that belong to the ATP-binding cassette (ABC) superfamily span from bacteria to humans [1–3]. These proteins seem to be associated with transport of solutes (from ions to proteins) across biological membranes. Examples of ABC proteins include molecules of medical importance such as the cystic fibrosis transmembrane conductance regulator (CFTR) and P-glycoprotein (multidrug resistance gene product; MDR1). These proteins are of particular interest because (1) mutations in CFTR, a Cl^- channel, cause cystic fibrosis, (2) the sulphonylurea receptor is directly involved in the response to oral hypoglycaemic agents used to treat type II diabetes, and (3) the expression of MDR1 in cancer cells confers multidrug resistance against chemotherapeutic agents [1–3]. One of the distinctive characteristics of ABC proteins is the presence of nucleotide-binding domains (NBDs) [1–3]. The proteins mentioned above contain two multispanning transmembrane domains, each followed by one NBD, and a linker region distal to the first NBD [1–3].

There is increasing interest in proteins of the ABC superfamily. In particular there is an interest in the NBDs, which are responsible for nucleotide binding and hydrolysis, and the linker region, which is involved in regulation of function and dimerization [1–6]. The linker region of CFTR is a target for phosphorylation mediated by protein kinase A (PKA) and protein kinase C (PKC), and is known as the regulatory domain [2]. The smaller similar linker region of MDR1 is known as minilinker domain (MLD; Figure 1A). Similarly to the the regulatory domain of CFTR, MDR1 MLD is phosphorylated by PKA and PKC, and its phosphorylation seems to be involved in the regulation of MDR1 [4,5]. Specifically, it has been claimed that MLD phosphorylation by PKA and PKC modulates swelling-activated Cl^- channels [4,6a].

In spite of recent advances in the knowledge of the structure and function of NBDs and linker regions with the use of peptides expressed in heterologous systems [2,3,6,7], the expression of ABC protein fragments in *Escherichia coli* has major problems. In fact, only detergent-extracted peptides and/or fusion proteins have been employed (i.e. fragments of interest not exposed to solubilizing agents have not been studied in isolation [8–13]). The availability of purified functional NBDs/linker regions in sufficient amounts for biochemical and structural studies would be very useful in the elucidation of the bases of ATP binding and hydrolysis by NBDs of ABC proteins and the mechanisms of protein regulation by the linker regions.

The aim of the present study was to identify an *E. coli* system for the expression of the N-terminal NBD (NBD1) of MDR1 followed by the MLD (NBD1MLD), with the aim of developing

Abbreviations used: ABC, ATP-binding cassette; CFTR, cystic fibrosis transmembrane conductance regulator; DTT, dithiothreitol; GST, glutathione S-transferase; IPTG, isopropyl β-D-thiogalactoside; MBP, maltose-binding protein; MDR1, human multidrug resistance gene product, human P-glycoprotein; MLD, MDR1 linker region, minilinker domain; NBD, nucleotide-binding domain; NBD1, N-terminal nucleotide-binding domain; NBD1MLD, MDR1 fragment containing NBD1 and MLD; NEM, *N*-ethylmaleimide; Ni-IDA, Ni^{2+}-imidodiacetic acid; PKA, protein kinase A; PKC, protein kinase C; TRX, thioredoxin.

[1] To whom correspondence should be addressed (e-mail Galtenbe@utmb.beach.edu).

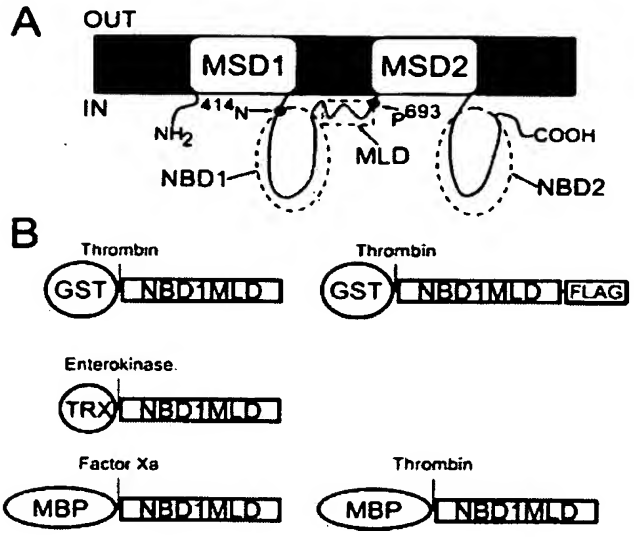


Figure 1 Fusion proteins expressed in *E. coli*

(**A**) Schematic representation of MDR1. Abbreviation: MSD, multiple spanning domain. The first (Glu^{414}) and last (Pro^{693}) residues of the NBD1MLD are labelled. (**B**) Schematic representation of the fusion proteins. Abbreviations: NBD2, C-terminal nucleotide-binding domain; FLAG, 8-residue FLAG epitope (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys). Specific proteolytic sites are indicated.

a system that yields sufficient soluble protein that can be easily purified for biochemical and structural studies. We therefore compared the expression and purification of MDR1 NBD1MLD fused to the C-terminus of glutathione S-transferase (GST), thioredoxin (TRX) and maltose-binding protein (MBP).

EXPERIMENTAL

E. coli strains and vectors

In all the experiments shown, we used the *E. coli* strain BL21 (lacking *ompT* and *lon* proteases). We also used JM109, DH5α and TOP10 strains in pilot experiments (see the Results section). The NBD1MLD fragment (280 amino acids; Asn^{414}–Pro^{693} of MDR1) was amplified from MDR1 cDNA (ATCC 65705) by PCR with the following primers: forward, 5′-CGGGATCCAACCTGAAGGTGCAGAGTGG-3′; reverse, 5′-CCGCTCGAGCTAAGGTATACTTTCATCCAGAG-3′. These primers contained *Bam*HI and *Xho*I sites (underlined) for cloning of the DNA fragment, in frame, into the fusion expression vectors. The expression vectors were pGEX-4t-3 (GST fusion vector; Pharmacia), pThioHis C (TRX fusion vector; InVitrogen) and pMal-c2 (MBP fusion vector; New England Biolabs). The GST, TRX and MBP fusion vectors contained thrombin, enterokinase and Factor Xa proteolytic sites respectively, to cleave the protein of interest from the fusion protein. The PCR fragment digested with *Bam*HI and *Xho*I was cloned into pGEX-4t-3, pThioHis C and pMal-c2 by using the *Bam*HI–*Xho*I, *Bgl*II–*Sal*I and *Bam*HI–*Sal*I sites respectively. The expression plasmids were named pGEX-NBD1MLD, pTRX-NBD1MLD and pMBP-NBD1MLD. The NBD1MLD DNA was also cloned into a modified pMal-c2 vector to which we added a thrombin site (underlined) by using the following oligonucleotide adaptors: forward, 5′-AATTCGGTACCCTGGTTCCGCGTG-3′; reverse, 5′-GATCCACGCGGAACCAGGGTACCG-3′. The annealed oligonucleotides, which contained a *Kpn*I site for easy primary screening, were ligated into the pMal-c2 vector cut with *Eco*RI and *Bam*HI. The NBD1MLD DNA cut with *Bam*HI and *Xho*I was cloned into the *Bam*HI and *Sal*I sites of the new expression vector (pMBPT); the resulting plasmid was named pMBPT-NBD1MLD. Finally, we also amplified NBD1MLD containing a FLAG epitope (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys) [14] at the C-terminus. We used the forward primer described above and the reverse primer 5′-CCGCTCGAGTCACTTGTCATCGTCGTCCTTGTAGTCAGGTATACTTTCATCCAGAGCCTC-3′. The PCR product was cut with *Bam*HI and *Xho*I, then cloned into the same sites of pGEX-4t-3, to express NBD1MLD fused to GST at the N-terminus and a FLAG epitope at the C-terminus (pGEX-NBD1MLDFLAG). Sequences of plasmid clones were confirmed by DNA sequencing. A schematic representation of all fusion proteins employed in the current studies is shown in Figure 1(B).

Expression of recombinant NBD1MLD

Unless stated otherwise, bacteria transformed with the expression vectors with or without NBD1MLD cDNA were grown in 100 ml of Luria–Bertani medium with 0.1 mg/ml ampicillin and 0.5% glucose to an A_{600} of 0.9–1.0 at 37 °C. Induction was performed with 0.1 mM isopropyl β-D-thiogalactoside (IPTG) at 37 °C; bacteria were harvested 2 h later by centrifugation at 4000 *g* for 15 min. The bacterial pellet was frozen at −80 °C in the specific binding buffers for the different fusion proteins, as follows: for GST fusion proteins, 150 mM NaCl/20 mM sodium phosphate/0.5 mM dithiothreitol (DTT)/10 mM EDTA (pH 7.3); for GST/FLAG fusion protein, 50 mM Tris/HCl/150 mM NaCl/0.5 mM DTT/10 mM EDTA (pH 7.4); for TRX fusion protein, 20 mM sodium phosphate/500 mM NaCl (pH 7.8); for MBP fusion proteins, 20 mM Tris/HCl/200 mM NaCl/1 mM EDTA (pH 7.4). PMSF from a stock in ethanol was added to the binding buffers at a concentration of 0.1–0.5 mM. The bacteria were thawed at 37 °C and lysed with a probe sonicator (three times at 150 W for 10 s, with 8 min intervals in a solid-CO_2/methanol bath). Part of the lysate was kept for SDS/PAGE analysis; the remainder was centrifuged at 265000 *g* for 15 min. The fusion proteins in the supernatant were then purified as described below. In some experiments the lysate was incubated with 1% (v/v) Triton X-100, then mixed continuously on a rotary mixer for 30 min. The resulting supernatant was collected by centrifugation as described above.

Purification of NBD1MLD hybrid proteins

Purification of GST fusion proteins was performed at room temperature. The soluble fractions containing GST fusion proteins were loaded on prepacked glutathione–Sepharose 4B columns (2 ml; Pharmacia), washed with 150 mM NaCl/20 mM sodium phosphate (pH 7.3) (10 column vol.) and eluted with 10 ml of 50 mM Tris/HCl, pH 8.0; 10 fractions of 1 ml were collected. The soluble fraction containing NBD1MLD with a FLAG epitope was loaded into a 1 ml column packed with FLAG M2 affinity gel (Eastman Kodak). The column was washed with 10–15 vol. of binding buffer; the fusion protein was eluted with 10 ml of 0.1 M glycine, pH 3. Each 1 ml fraction collected was immediately neutralized with 20 μl of 1 M Tris base, pH 8.0. The soluble fraction containing the TRX fusion protein was loaded on a 3 ml column packed with a Ni^{2+}-imidodiacetic acid (Ni-IDA) affinity gel (ProBond; InVitrogen). The column was washed with 8 ml of binding buffer, then with

20 mM sodium phosphate/500 mM NaCl (pH 6.0), until the eluate A_{280} was less than 0.01. Then the elution proceeded with 20 mM sodium phosphate/500 mM NaCl (pH 6.0) containing increasing concentrations of imidazole in sequential steps. The concentrations of imidazole were 50, 200, 350 and 500 mM, 5 ml each step; fractions of 1 ml were collected. All the steps for TRX–NBD1MLD purification were performed at 4 °C. Purification of MBP fusion proteins was also conducted at 4 °C. The soluble fractions containing the MBP fusion proteins were loaded into a 1 cm × 10 cm column packed with amylose affinity gel (New England Biolabs) and washed with 10–15 column vol. of the binding buffer. Elution was performed with 15 ml of the same buffer containing 10 mM maltose. Fractions of 3 ml were collected.

Purification of NBD1MLD

Purified MBPT–NBD1MLD in the elution buffer was digested with thrombin (10 units/mg of fusion protein; Calbiochem, La Jolla, CA, U.S.A.) for 2 h at room temperature. At the end of the digestion period, 1 mM benzamidine was added to inactivate the thrombin. The reaction mix was desalted on a PD-10 column (Pharmacia) equilibrated with 20 mM Tris/HCl/100 mM NaCl/10 mM DTT (pH 7.0) and purified on a 3 ml DEAE-Sephacel column (Sigma, St. Louis, MO, U.S.A.). After washing with 15 vol. of the same buffer, elution was performed by lowering the buffer pH to 6.5 and increasing the NaCl concentration to 0.5 M. The fractions containing NBD1MLD were pooled and desalted. NBD1MLD was finally purified by gel filtration after exchange of the buffer to 50 mM Tris/HCl/150 mM NaCl/0.1 mM EDTA/1 mM DTT (pH 7.4). The DEAE-purified sample was subjected to gel filtration on an FPLC system (Pharmacia). The sample was loaded on a 1 cm × 30 cm column packed with Superdex-75 or Superdex-200 (Pharmacia) pre-equilibrated with the sample buffer. Elution was at a flow rate of 0.4 ml/min (with the same buffer). Fractions of 0.8 ml were collected and analysed for the presence of NBD1MLD by staining with Coomassie Blue after PAGE.

Determination of ATPase activity

ATPase activity was calculated from P_i release measured colorimetrically [15]. Determinations were performed in 125 μl of 50 mM Tris/HCl/0.1 mM EDTA/1 mM DTT (pH 8.0) in the presence of 5 mM ATP and 2.5 mM $MgCl_2$. Reactions were initiated by addition of the $MgCl_2$, and proceeded for 10 min at 37 °C. No ATP hydrolysis was detected in the absence of $MgCl_2$ or protein; P_i release was constant during the reaction.

Other techniques

SDS/PAGE, Western-blotting analysis and protein determinations were performed with standard techniques, as previously described [16,17]. The C219 monoclonal antibody (Signet, Dedham, MA, U.S.A.) recognizes an epitope in MDR1 NBD [18]. Densitometric analysis of Coomassie Blue-stained gels were performed with SigmaGel (Jandel). Peptide sequencing was performed by J. S. Smith (Protein Chemistry Laboratory, University of Texas Medical Branch, Galveston, TX, U.S.A.).

RESULTS AND DISCUSSION

In spite of the recent advances in the knowledge of the structure and function of NBDs and linker regions expressed in heterologous systems, the expression of ABC protein fragments in *E. coli* has been problematic. First, the expression of NBDs of ABC proteins and the regulatory domain of CFTR is directed to inclusion bodies [8–11]. This problem has been solved in part by expression of the fragments as fusion proteins. However, some of the fusion peptides are still directed exclusively to inclusion bodies [8–11]. Secondly, on many occasions, fusion proteins are not recovered as soluble proteins in the absence of Triton X-100 [8,12,13], and it is likely that detergent extraction disrupts NBD function [19]. Finally, the expression of the linker regions, either by themselves or following the N-terminal NBD (NBD1), is toxic and considerably decreases the expression in *E. coli* [20]. Therefore there are many aspects of the expression/purification of NBDs/linker regions of ABC proteins that need to be improved to obtain preparations suitable for detailed biochemical and structural studies.

Because it is of interest to express the first NBD of ABC proteins with the linker region to establish whether this region regulates NBD function [21], we decided to identify an adequate expression system in *E. coli* suitable for this task. Because of our primary interest in MDR1, we wished to optimize the expression and purification of NBD1MLD, but our results could be applicable to equivalent fragments of other ABC proteins as well as to poorly soluble proteins in general. Our results also provide comparative information on the expression of GST, TRX and MBP fusion proteins that might be useful to researchers interested in protein expression in *E. coli*.

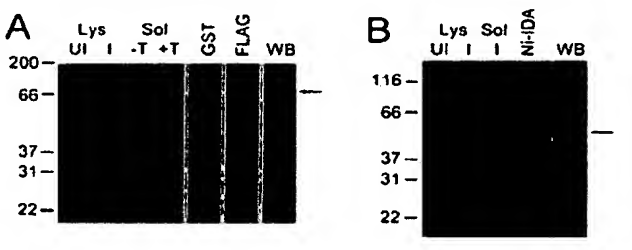


Figure 2 Expression and purification of GST and TRX fusion proteins

Analysis of expression and purification of fusion proteins by SDS/PAGE. Proteins were stained with Coomassie Blue. **(A)** Expression and purification of GST fusion proteins. Lanes show protein electrophoresis of cell lysate (Lys) from cells induced with IPTG (I) or uninduced (UI), soluble fraction (Sol) without (−T) or with (+T) Triton X-100 extraction, and purification by glutathione–Sepharose 4B (GST) and M2-gel affinity chromatography (FLAG). Results from BL21 cells transformed with pGEX-NBD1MLDFLAG are shown in all cases except in the lane labelled GST, for which cells were transformed with pGEX-NBD1MLD (no FLAG epitope at the C-terminus). The lane labelled WB corresponds to a Western blot of purified GST–NBD1MLD. **(B)** Expression and purification of TRX–NBD1MLD. Lane labelling is as described for **(A)** except for Ni-IDA, which denotes purification by agarose–Ni-IDA affinity chromatography. The lane labelled WB is an immunoblot of the Ni-IDA affinity purification. The protein seen in the Ni-IDA lane is not TRX–NBD1MLD, but a co-purified protein that was not recognized by the anti-(P-glycoprotein) monoclonal antibody C219 in immunoblots (the protein was present in elution fractions without NBD1MLD; results not shown). Soluble TRX–NBD1MLD was never detected in gels stained with Coomassie Blue. The anti-(P-glycoprotein) antibody C219 was employed for all immunoblots. The equivalent of 0.3 ml of *E. coli* growth was loaded in the Lys lanes. The Sol lanes contained 250 μg of protein. The amount of protein in lanes GST, FLAG and WB of **(A)** was 0.3 μg. The amount of protein in the Ni-IDA and WB lanes of **(B)** was 0.5 μg. The arrows indicate the position of the fusion proteins.

Expression of NBD1MLD fused to GST or TRX as soluble proteins is low

Figure 2 shows that the GST, with or without (results not shown) a FLAG epitope, and the TRX fusion expression systems did not yield high levels of expression of soluble NBD1MLD fusion

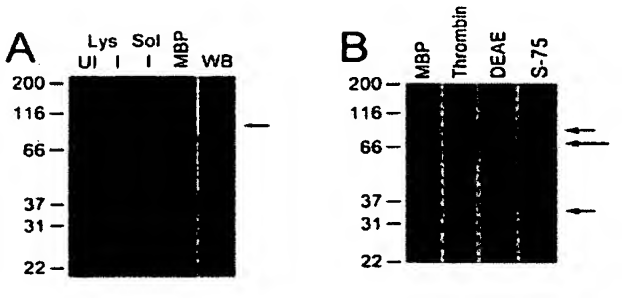


Figure 3 Expression and purification of NBD1MLD fused to MBP

(**A**) Expression and purification of MBP–NBD1MLD. Lane labelling is as described in the legend to Figure 2 except for MBP, which refers to purification by amylose-affinity chromatography. The arrow indicates the position of the fusion protein monomer. The C219 antibody was used for the Western blot (WB lane). (**B**) Analysis of the purification of NBD1MLD from the MBP fusion protein. Proteins were analysed by SDS/PAGE and stained with Coomassie Blue. Lane MBP, purification of NBD1MLD fused to MBP by amylose-affinity chromatography; lane Thrombin, thrombin cleavage of NBD1MLD from the fusion protein; lane DEAE, removal of MBP by DEAE anion-exchange chromatography; lane S-75, NBD1MLD purification by gel filtration on Superdex-75. The small arrows point to the MBP fusion protein and NBD1MLD. The equivalent of 0.3 ml of *E. coli* growth was loaded in the Lys lanes. The Sol lane contained 250 μg of protein. The amount of protein in lanes MBP and WB of (**A**) was 0.5 μg. The amount of protein in the MBP, thrombin and S-75 lanes of (**B**) was 6–7 μg. The DEAE lane contained approx. 1 μg of protein. The large arrow points to the NBD1MLD dimer subjected to electrophoresis after gel filtration.

proteins. A FLAG epitope at the C-terminus of GST–NBD1-MLD can be used for the chromatographic separation of GST–NBD1MLD from GST or truncated proteins containing GST, without compromising the expression (compare GST and FLAG lanes in Figure 2A). The expression of soluble or Triton X-100-extracted GST fusion proteins, as well as that of soluble TRX–NBD1MLD, was not affected significantly by heat shock of the bacteria immediately before induction, the use of several different *E. coli* strains, lowering the temperature after IPTG induction, varying the IPTG concentration, induction at lower bacterial density, or increasing the induction time to 3.5 h (results not shown). Our results are consistent with the observations that expression of NBDs as fusion proteins with polyhistidine or GST did not improve solubility (see Figure 2A) [8–12]. Although we do not know the reasons for the low expression of soluble NBD/linker-region-containing proteins, it might be related to their hydrophobicity. It is known that the overexpression of membrane proteins is difficult [24] and that NBDs of CFTR and histidine permease do solubilize in lipid membranes [25–28].

Expression of NBD1MLD fused to MBP as a soluble protein is significant, and functional NBD1MLD can be purified from the fusion protein by using a vector with an added thrombin site

Figure 3(A) illustrates the expression of MBP–NBD1MLD in BL21 cells. Although more than 70% of the fusion protein was directed to inclusion bodies, a significant amount (approx. 8% of *E. coli* protein) was expressed as soluble protein (Figure 3A). Moreover, the fusion protein could be purified in a single step on the basis of its binding to amylose (Figure 3A). Formation of the high-molecular-mass aggregates of MBP–NBD1MLD seen in Figure 3(A) could be prevented by 1 mM DTT (compare the MBP lanes in Figures 3A and 3B), suggesting a role of disulphide bonds from the Cys in motif A of NBD (the only Cys in MBP–NBD1MLD; Cys^{431} of MDR1).

Unfortunately, cleavage of NBD1MLD from the fusion protein using the Factor Xa site could not be optimized, because we always observed the proteolysis of NBD1MLD. Inefficient and non-specific digestion is a common observation when using Factor Xa [30–32]. These results are in agreement with studies showing the high-level expression of NBDs of ABC proteins fused to MBP, which could not be cleaved off efficiently with Factor Xa [30,31]. Because we could obtain adequate digestion of NBD1MLD from GST–NBD1MLD by using thrombin, we modified the commercial MBP fusion protein vector pMal-c2 by adding a thrombin site. The resulting fusion protein containing the added thrombin site between MBP and NBD1MLD (MB-PT–NBD1MLD) was expressed at a level undistinguishable from that of MBP–NBD1MLD. Furthermore NBD1MLD could be cleaved off from MBP and purified by a combination of anion-exchange and gel-filtration chromatographies (see Figure 3B). MBP did not bind to DEAE and could be removed easily from the thrombin digestion reaction. The DEAE-purified material was approx. 80% NBD1MLD [two proteins more than 30 kDa, barely seen in the DEAE lane of Figure 3(B), were enriched]. A highly purified (more than 95%) NBD1MLD could be collected in the void volume of the Superdex-75 gel-filtration column (also in the void volume when Superdex-200 was used). Formation of aggregates is a common feature of NBD peptides [30].

The identity of MBP–NBD1MLD was confirmed by the presence of MBP (the fusion protein can be purified on an amylose column; see Figure 3), the apparent molecular mass of the fusion protein (approx. 78 kDa; Figure 3) and the thrombin cleavage products (approx. 42 and 36 kDa; Figure 3B), the presence of the C219 epitope (near the end of NBD1; see Figure 3A), and the N-terminal sequence (Gly-Ser-Asn-Leu-Lys) of NBD1MLD after the digestion of MBP–NBD1MLD by thrombin.

Finally, we measured the ATPase activities of purified MB-PT–NBD1MLD, DEAE-purified NBD1MLD and purified NBD1MLD. All peptides had ATPase activities comparable with those of other NBDs of ABC proteins. The ATPase activity of MBP–NBD1MLD was inhibited by 47 ± 3% (mean $\pm$ S.E.M., $n = 4$) by 500 μM of *N*-ethylmaleimide (NEM). NEM is a thiol reagent known to block P-glycoprotein ATPase activity [33]. Interestingly, the inhibition of the ATPase activity by NEM was abolished when 5 mM ATP was present before the addition of NEM (4 ± 2% inhibition as compared with control; $n = 3$). Protection by ATP against NEM inhibition has been described for full-length P-glycoprotein [33]. These results indicate that the NBD1MLD is functional and that the anion-exchange and gel-filtration purification procedures do not decrease the ATPase activity of NBD1MLD (see Table 1).

Table 1 Purification and ATPase activity of MBP–NBD1MLD and NBD1MLD

Total protein refers to the protein amount/litre of bacterial growth. Approximate purity was calculated from densitometry of Coomassie Blue-stained gels. The ATPase activity is the mean $\pm$ S.E.M. and the number of determinations is shown in parentheses. Abbreviation: n.d., not determined. See the Experimental section for details.

Preparation	Total protein (mg/l of culture)	ATPase activity (nmol of P_i/min per mg)	Purity (%)
Soluble fraction	580	n.d.	12
Amylose affinity	70	57 ± 2 (5)	> 95
DEAE chromatography	18	92 ± 2 (3)	> 80
Gel filtration	3.5	144 ± 4 (3)	> 95

Summary

Comparison of NBD1MLD expression as fusion proteins with GST, with TRX and with MBP shows that a high level of expression in the soluble fraction (approx. 8% of total *E. coli* protein) can be achieved only for MBP–NBD1MLD (see Table 1). The addition of a proteolytic thrombin site just proximal to the N-terminal end of NBD1MLD permits the cleavage of NBD1MLD from MBP. NBD1MLD can be purified easily by ion-exchange and gel-filtration chromatographies with retention of its ATPase activity. The approach described here could be generally applicable to solving the problems of expression and purification of NBDs/linker regions of ABC proteins.

We thank Dr. S. King and Dr. L. Reuss for comments on a preliminary version of the manuscript, and J. S. Smith (Protein Chemistry Laboratory, University of Texas Medical Branch, Galveston, TX, U.S.A.) for the peptide sequencing. This work was supported in part by the American Heart Association, Texas Affiliate Grant 966-1613, Searle Research and Development, and National Cancer Institute Grant CA72783.

REFERENCES

1 Philipson, L. H. and Steiner, D. F. (1995) Science **268**, 372–373
2 Riordan, J. R. (1993) Annu. Rev. Physiol. **55**, 609–630
3 van Veen, H. W. and Konings, W. N. (1997) Cancer Biol. **8**, 183–191
4 Hardy, S. P., Goodfellow, H. R., Valverde, M. A., Gill, D. R., Sepulveda, F. V. and Higgins, C. F. (1995) EMBO J. **14**, 68–75
5 Juvvadi, S. R., Glavy, J. S., Horwitz, S. B. and Orr, G. A. (1997) Biochem. Biophys. Res. Commun. **230**, 442–447
6 Winter, M. C. and Welsh, M. J. (1997) Nature (London) **389**, 294–296
6a Vanoye, C. G., Castro, A. F., Reuss, L and Altenberg, G. A. (1999) Am. J. Physiol., in the press
7 Dulhanty, A. M. and Riordan, J. R. (1994) Biochemistry **33**, 4072–4079
8 Dayan, G., Baubichon-Cortay, H., Jault, J. M., Cortay, J. C., Deleage, G. and Di Pietro, A. (1996) J. Biol. Chem. **271**, 11652–11658
9 Hartman, J., Huang, J., Rado, T., Peng, S., Jilling, T., Muccio, D. and Sorscher, E. (1992) J. Biol. Chem. **267**, 6455–6458
10 Muller, K. M., Ebensperger, C. and Tampe, R. (1994) J. Biol. Chem. **269**, 14032–14037
11 Yike, I., Ye, J., Zhang, Y., Manavalan, P., Gerken, T. A. and Dearborn, D. G. (1996) Protein Sci. **5**, 89–97
12 Baubichon-Cortay, H., Baggetto, L. G., Dayan, G. and Di Pietro, A. (1994) J. Biol. Chem. **269**, 22983–22989
13 Randak, C., Roscher, A. A., Hadorn, H.-B., Assfalg-Machleidt, I., Auerswald, E. A. and Machleidt, W. (1995) FEBS Lett. **363**, 189–194
14 Hopp, T. P., Prickett, K. S., Price, V., Libby, R. T., Cerretti, P., Urdal, D. L. and Conlon, P. J. (1988) Bio/Technology **6**, 1205–1210
15 Chifflet, S., Torriglia, A., Chiesa, R. and Tolosa, S. (1988) Anal. Biochem. **168**, 1–4
16 Altenberg, G. A., Subramanyam, M. and Reuss, L. (1994) Am. J. Physiol. **267**, C1196–C1202
17 Castro, A. F. and Altenberg, G. A. (1997) Biochem. Pharmacol. **53**, 89–93
18 Georges, E., Bradley, G., Gariepy, J. and Ling, V. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 152–156
19 Randak, C., Neth, P., Auerswald, E. A., Eckerskorn, C., Assfalg-Machleidt, I. and Machleidt, W. (1997) FEBS Lett. **410**, 180–186
20 Yike, I., Zhang, Y., Ye, J. and Dearborn, D. G. (1996) Protein Expression Purif. **7**, 45–50
21 Neville, D. C. A., Rozanas, C. R., Tulk, B. M., Townsend, R. R. and Verkman, A. S. (1998) Biochemistry **37**, 2401–2409
22 Reference deleted
23 Reference deleted
24 Grisshammer, R. and Tate, C. G. (1995) Quart. J. Biophys. **8**, 315–422
25 Arispe, N., Rojas, E., Hartman, J., Sorscher, E. J. and Pollard, H. B. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 1539–1543
26 Baichwal, V., Liu, D. and Ames, G. F. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 620–624
27 Gruis, D. B. and Price, E. M. (1997) Biochemistry **36**, 7739–7745
28 Ko, Y. H., Delanoy, M. and Pedersen, P. L. (1997) Biochemistry **36**, 5053–5064
29 Reference deleted
30 Ko, Y. H., Thomas, P. J., Delannoy, M. R. and Pedersen, P. L. (1993) J. Biol. Chem. **268**, 24330–24338
31 Sharma, S. and Rose, D. R. (1995) J. Biol. Chem. **270**, 14085–14093
32 Wearne, S. J. (1990) FEBS Lett. **263**, 23–26
33 Al-Shawi, M. K., Urbatsch, I. L. and Senior, A. E. (1994) J. Biol. Chem. **269**, 8986–8992

Received 18 August 1998/2 November 1998; accepted 24 November 1998